

"EXPRESS MAIL" MAILING LABEL NUMBER TT3376467447 US

DATE OF DEPOSIT JUNE 7, 1995

I HEREBY CERTIFY THAT THIS PAPER OR FEE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE "EXPRESS MAIL POST OFFICE TO ADDRESSEE" SERVICE UNDER 37 CFR 1-10 ON THE DATE INDICATED ABOVE AND IS ADDRESSED TO THE COMMISSIONER OF PATENTS AND TRADEMARKS, WASHINGTON, D.C. 20231.

GREBB LOUGAY
(TYPED OR PRINTED NAME OF PERSON MAILING PAPER OR FEE)

[Signature]
(SIGNATURE OF PERSON MAILING PAPER OR FEE)

A P P L I C A T I O N

for

UNITED STATES LETTERS PATENT

on

CLONING AND RECOMBINANT PRODUCTION
OF RECEPTOR(S) OF THE ACTIVIN/TGF- β SUPERFAMILY

by

Lawrence S. Mathews,

Wylie W. Vale,

and

Kunihiro Tsuchida

Number of Drawings: Six

Docket No.: P41 9981

Salk File No.: S95045

Attorneys

Pretty, Schroeder, Brueggemann & Clark
444 South Flower Street, Suite 2000
Los Angeles, California 90071

CLONING AND RECOMBINANT PRODUCTION
OF RECEPTOR(S) OF THE ACTIVIN/TGF- β SUPERFAMILY

RELATED APPLICATIONS

This application is a continuation-in-part of United States Serial Number 08/300,584, filed September 2, 1994, now pending, which is a continuation of United States Serial Number 07/880,220, filed May 8, 1992, now abandoned, which is a continuation-in-part of United States Serial Number 07/773,229, filed October 9, 1991, now abandoned, which is, in turn, a continuation-in-part of United States Serial Number 07/698,709, filed May 10, 1991, now abandoned.

ACKNOWLEDGEMENT

This invention was made with Government support under Grant Numbers HD 13527 and DK 26741, awarded by the National Institutes of Health. The Government has certain rights in this invention.

FIELD OF THE INVENTION

The present invention relates to receptor proteins, DNA sequences encoding same, and various uses therefor.

BACKGROUND OF THE INVENTION

Activins are dimeric proteins which have the ability to stimulate the production of follicle stimulating hormone (FSH) by the pituitary gland. Activins share a common subunit with inhibins, which inhibit FSH secretion.

Activins are members of a superfamily of polypeptide growth factors which includes the inhibins, the transforming growth factors- β (TGF- β), Mullerian duct

inhibiting substance, the *Drosophila* decapentaplegic peptide, several bone morphogenetic proteins, and the Vg-related peptides.

5 As a result of their extensive anatomical distribution and multiple biological actions, members of this superfamily of polypeptide growth factors are believed to be involved in the regulation of numerous biological processes. Activin, for example, is involved in the
10 proliferation of many tumor cell lines, the control of secretion and expression of the anterior pituitary hormones (e.g., FSH, GH and ACTH), neuron survival, hypothalamic oxytocin secretion, erythropoiesis, placental and gonadal steroidogenesis, early embryonic development, and the like.

15 Other members of the activin/TGF- β superfamily of polypeptide growth factors are involved in the regulation of cell function and cell proliferation for numerous cell types, in adults and embryos. For example, cells which are
20 subject to regulation by one or more members of the activin/TGF- β superfamily of polypeptide growth factors include mesenchymal cells, muscle cells, skeletal cells, immune cells, hematopoietic cells, steroidogenic cells, endothelial cells, liver cells, epithelial cells, and the
25 like.

Chemical cross-linking studies with a number of cell types suggests that multiple binding sites (i.e., receptors) exist on the surface of cells. However, little
30 is known about the structure of these receptors, or about the second messenger signalling systems that they employ. It would be desirable, therefore, if the nature of these poorly characterized receptor proteins could be more fully understood.

35

BRIEF DESCRIPTION OF THE INVENTION

In accordance with the present invention, we have identified and characterized members of a new superfamily of receptor proteins which comprise three distinct domains: an extracellular, ligand-binding domain, a hydrophobic, trans-membrane domain, and an intracellular, receptor domain having serine kinase-like activity.

Also provided are DNAs encoding the above-described receptor proteins, and antibodies thereto, as well as bioassays, therapeutic compositions containing such proteins and/or antibodies, and applications thereof.

The DNAs of the invention are useful as probes for the identification of additional members of the invention superfamily of receptor proteins, and as coding sequences which can be used for the recombinant expression of the invention receptor proteins, or functional fragments thereof. The invention receptor proteins, and antibodies thereto, are useful for the diagnosis and therapeutic management of carcinogenesis, wound healing, disorders of the immune, reproductive, or central nervous systems, and the like.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a schematic diagram of receptors of the invention and the various domains thereof.

Figure 2 outlines the strategy used for expression cloning of a receptor of the activin/TGF- β receptor superfamily.

Figure 3 is a schematic of two mouse activin receptor clones. The top line of the figure is a restriction map, in kb, of mActR1 and mActR2, with

numbering starting from bp 1 of mActR2. The dotted line in the figure represents 5' untranslated sequences present only in mActR1. The middle lines present a schematic representation of two activin receptor cDNA clones. Boxes represent coding sequences---black is the signal peptide, white is the extracellular ligand-binding domain, gray is the transmembrane, and the intracellular kinase domain is hatched. Amino acids are numbered beneath the schematics.

Figure 4 presents a comparison between activin receptor and daf-1 [a *C. elegans* gene encoding a putative receptor protein kinase (with unknown ligand); see Georgi, et al., *Cell* 61: 635-645 (1990)]. Conserved residues between the activin receptor and daf-1 are highlighted; conserved kinase domain residues are designated with an "**".

Figure 5A summarizes results of ^{125}I activin A binding to COS cells transfected with pmActR1. Binding was competed with unlabeled activin A. For the runs reported herein, total binding was 4.6% of input cpm, non-specific binding was 0.9% of input cpm, and therefore the specific binding was 3.7% of input cpm. Data are shown as % specific binding, normalized to 100%. The inset presents a Scatchard analysis of the data [*Ann. NY Acad. Sci.* 51: 660-672 (1979)].

Figure 5B summarizes results of ^{125}I activin A binding to COS cells transfected with pmActR2. Binding was competed with unlabeled factors as indicated in the figure. For the runs reported herein, total binding was 3.4% of input cpm, non-specific binding was 0.9% of input cpm, and therefore the specific binding was 2.5% of input cpm. Data are shown as % specific binding, normalized to 100%.

Figure 6 is a phylogenetic tree, comparing the relationship of the activin receptor kinase domain to other

protein kinases. To construct the tree, the catalytic domains of representative sequences were empirically aligned and evolutionary relatedness was calculated using an algorithm designed by Fitch and Margoliash [Science 155: 279-284 (1967)], as implemented by Feng and Doolittle [J. Mol. Evol. 25: 351-360 (1987)]. Known subfamilies of kinases are indicated in the figure. For those sequences that had similarity scores (i.e., a relative sequence identity) of at least 4 standard deviations above the mean (in comparison with all other known kinase sequences), the percent identity with the activin receptor is indicated. For further detail on kinase sequences, the reader is referred to Hanks and Quinn, Meth. Enzymol. 200: 38-62 (1991).

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the present invention, there is provided a novel superfamily of receptor protein(s) characterized by having the following domains, reading from the N-terminal end of said protein:

an extracellular, ligand-binding domain,
a hydrophobic, trans-membrane domain, and
an intracellular domain having serine kinase-like activity.

The novel receptor protein(s) of the invention optionally further comprise a second hydrophobic domain at the amino terminus thereof.

As employed herein, the phrase "extracellular, ligand-binding domain" refers to that portion of receptors of the invention which has a high affinity for ligand, and which, when associated with a cell, resides primarily outside of the cell membrane. Because of its location, this domain is not exposed to the processing machinery present within the cell, but is exposed to all components

of the extracellular medium. See Figure 1.

As employed herein, the phrase "hydrophobic, trans-membrane domain" refers to that portion of receptors of the invention which traverses the cell membrane, and serves as a "bridge" between the extracellular and intracellular domains of the receptor. The hydrophobic nature of this domain serves to anchor the receptor to the cell membrane. See Figure 1.

10

As employed herein, the phrase "intracellular domain having serine kinase-like activity" refers to that portion of receptors of the invention which resides within the cytoplasm, and which embodies the catalytic functionality characteristic of all receptors of the invention. See Fig 1.

The optional second hydrophobic domain, positioned at the amino terminus of receptors of the invention, comprises a secretion signal sequence which promotes the intracellular transport of the initially expressed receptor protein across the Golgi membrane. See Figure 1.

Members of the invention superfamily of receptors can be further characterized as having sufficient binding affinity for at least one member of the activin/TGF- β superfamily of polypeptide growth factors such that concentrations of ≤ 10 nM of said polypeptide growth factor occupy $\geq 50\%$ of the binding sites of said receptor protein.

Binding affinity (which can be expressed in terms of association constants, K_a , or dissociation constants, K_d) refers to the strength of interaction between ligand and receptor, and can be expressed in terms of the concentration of ligand necessary to occupy one-half (50%) of the binding sites of the receptor. A receptor having a

high binding affinity for a given ligand will require the presence of very little ligand to become at least 50% bound (hence the K_d value will be a small number); conversely, receptor having a low binding affinity for a given ligand will require the presence of high levels of ligand to become 50% bound (hence the K_d value will be a large number).

Reference to receptor protein "having sufficient binding affinity such that concentrations of said polypeptide growth factor less than or equal to 10 nM (i.e., ≤ 10 nM) occupy $\geq 50\%$ (i.e., greater than or equal to one-half) of the binding sites of said receptor protein" means that ligand (i.e., polypeptide growth factor) concentration(s) of no greater than about 10 nM are required in order for the ligand to occupy at least 50% of the active sites of said receptor (preferably about 0.1-1.0 nM of said receptor), with much lower ligand concentrations typically being required. Presently preferred receptors of the present invention have a binding affinity such that ligand concentration(s) in the range of only about 100 - 500 pM are required in order to occupy (or bind to) at least 50% of the receptor binding sites, wherein the receptor concentration is preferably about 0.1-1.0 nM.

Members of the invention superfamily of receptors can be divided into various subclasses, based on the approximate size of the crosslinked complexes obtained when radiolabeled activin is chemically crosslinked to cell extracts [see, for example, Example VI below, or Mathews and Vale in Cell 65:973-982 (1991)]. Type I activin/TGF- β receptors are those which form a crosslinked complex of about 65 kD with activin; Type II receptors are those which form a crosslinked complex of about 80-85 kD with activin; while Type III, Type IV and the like receptors are those which form crosslinked complexes with activin having molecular weights greater than about 100 kD.

Each member of a given subclass is related to other members of the same subclass by the high degree of homology (e.g., >80% overall amino acid homology; frequently having >90% overall amino acid homology) between
5 such receptors; whereas members of a given subclass differ from members of a different subclass by the lower degree of homology (e.g., at least about 30% up to 80% overall amino acid homology; with in the range of about 40% up to 90% amino acid homology specifically in the kinase domains
10 thereof) between such receptors. Typically, related receptors have at least 50% overall amino acid homology; with at least about 60% amino acid homology in the kinase domains thereof. Preferably, related receptors are defined as those which have at least 60% overall amino acid
15 homology; with at least about 70% amino acid homology in the kinase domains thereof.

Based on the above criteria, the receptors described herein are designated Type II receptors, with the
20 first discovered Type II receptor (i.e., the mouse-derived activin receptor) being designated ActRII, while subsequently identified Type II receptors which are not homologs of ActRII (because while clearly related by size and some sequence homology, they differ sufficiently to be
25 considered as variants of ActRII), are designated ActRIIB, ActRIIC, etc.

Presently preferred members of the invention superfamily of receptors are further characterized by
30 having a greater binding affinity for activins than for inhibins. Such receptors are frequently also observed to have:

substantially no binding affinity for transforming growth factors- β , and

35 substantially no binding affinity for non-activin-like proteins or compounds.

00742604 121000

5 Additional members of the invention superfamily of receptors are further characterized by having a greater binding affinity for TGF- β s than for activins or inhibins.

Transforming growth factors- β (TGF- β s) are members of the activin/TGF- β superfamily of polypeptide growth factors. TGF- β s are structurally related to activins, sharing at least 20-30% amino acid sequence homology therewith. TGF- β s and activins have a substantially similar distribution pattern of cysteine residues (or substitution) throughout the peptide chain. Furthermore, both polypeptides, in their active forms, are dimeric species.

As employed herein, the term "non-activin-like" proteins refers to any protein having essentially no structural similarity with activins (as defined broadly herein).

Preferred members of the invention superfamily of receptors comprise those having in the range of about 500 amino acids, and are further characterized by having the following designated sizes for each of the domains thereof,

5 reading from the N-terminal end of said receptor:

the extracellular, ligand-binding domain preferably will have in the range of about 88-118 amino acids,

10 the hydrophobic, trans-membrane domain preferably will have in the range of about 23-28 amino acids, beginning at the carboxy terminus of the extracellular domain, and

15 the intracellular domain having kinase-like activity preferably will have in the range of about 345-360 amino acids, beginning at the carboxy terminus of the hydrophobic, trans-membrane domain.

Receptors of the invention optionally further comprise a second hydrophobic domain having in the range of
20 about 16-30 amino acids at the extreme amino terminus thereof (i.e., at the amino terminus of the extracellular, ligand-binding domain). This domain is a secretion signal sequence, which aids the transport of invention receptor(s) across the cell membrane. Exemplary secretion signal
25 sequences include amino acids 1-19 of Sequence ID No. 1, amino acids 1-20 of Sequence ID No. 3, amino acids 1-25 of Sequence ID No. 11, and the like. Such secretion signal sequences can be encoded by such nucleic acid sequences as nucleotides 71-127 of Sequence ID No. 1, nucleotides 468-
30 527 of Sequence ID No. 3, nucleotides 72-146 of Sequence ID No. 11, and the like.

Members of the invention superfamily of receptors can be obtained from a variety of sources, such as, for
35 example, pituitary cells, placental cells, hematopoietic cells, brain cells, gonadal cells, liver cells, bone cells, muscle cells, endothelial cells, epithelial cells,

mesenchymal cells, kidney cells, and the like. Such cells can be derived from a variety of organisms, such as, for example, human, mouse, rat, ovine, bovine, porcine, frog, chicken, fish, mink, and the like.

5

Presently preferred amino acid sequences encoding receptor proteins of the invention include the sequence set forth in Sequence ID No. 2 (which represents a mouse activin receptor amino acid sequence), a modified form of Sequence ID No. 2 wherein the arginine at residue number 39 is replaced by a lysine, the isoleucine at residue number 92 is replaced by a valine, and the glutamic acid at residue number 288 is replaced by a glutamine (which modified form of Sequence ID No. 1 is referred to hereinafter as "Sequence ID No. 1'", and represents a human activin receptor amino acid sequence), the sequence set forth as Sequence ID No. 4 (which represents a *Xenopus* activin receptor amino acid sequence), and Sequence ID No. 12 (which represents a rat activin receptor-like kinase amino acid sequence) as well as functional, modified forms thereof. Those of skill in the art recognize that numerous residues of the above-described sequences can be substituted with other, chemically, sterically and/or electronically similar residues without substantially altering the biological activity of the resulting receptor species.

In accordance with another embodiment of the present invention, there is provided a soluble, extracellular, ligand-binding protein, further characterized by:

having sufficient binding affinity for at least one member of the activin/TGF- β superfamily of polypeptide growth factors such that concentrations of ≤ 10 nM of said polypeptide growth factor occupy $\geq 50\%$ of the binding sites on said receptor protein, and

having at least about 30% sequence identity with

respect to:

the sequence of amino acids 20-134 set forth in Sequence ID No. 2;

5 the sequence of amino acids 20-134 set forth in Sequence ID No. 2, wherein the arginine residue at position number 39 is replaced by a lysine, and the isoleucine at residue number 92 is replaced by a valine;

10 the sequence of amino acids 21-132 set forth in Sequence ID No. 4; or

the sequence of amino acids 26-113 set forth in Sequence ID No. 12.

15 Presently preferred soluble, extracellular, ligand-binding proteins contemplated by the present invention can be further characterized by having at least about 50% sequence identity with respect to:

the sequence of amino acids 20-134 set forth in Sequence ID No. 2;

20 the sequence of amino acids 20-134 set forth in Sequence ID No. 2, wherein the arginine residue at position number 39 is replaced by a lysine, and the isoleucine at residue number 92 is replaced by a valine;

25 the sequence of amino acids 21-132 set forth in Sequence ID No. 4; or

the sequence of amino acids 26-113 set forth in Sequence ID No. 12;

30 with the presently most preferred soluble, extracellular, ligand-binding proteins having at least about 80% sequence identity with respect to the above-referenced fragments of Sequence ID Nos. 2, 4 or 12.

35 Members of the class of soluble, ligand-binding proteins contemplated by the present invention may be divided into various subclasses, as previously described, wherein members of one subclass may have a greater binding

affinity for activins than for inhibins and/or TGF- β s; or alternatively, members of another subclass may have a greater binding affinity for inhibins than for activins and/or TGF- β s; or alternatively, members of yet another subclass may have a greater binding affinity for TGF- β s than for activins and/or inhibins. It is, of course, understood by those of skill in the art, that members of more than one subclass may have a greater binding affinity for one member of the activin/TGF- β superfamily of polypeptide growth factors, relative to other members of the superfamily.

Presently preferred soluble, extracellular, ligand-binding proteins of the present invention are further characterized by:

- having a greater binding affinity for activins than for inhibins,
- having substantially no binding affinity for transforming growth factors- β , and
- having substantially no binding affinity for non-activin-like proteins.

Presently preferred soluble, extracellular, ligand-binding proteins of the present invention typically comprise in the range of about 88-118 amino acids.

Especially preferred soluble, extracellular, ligand-binding proteins of the invention are those having substantially the same amino acid sequence as that set forth as:

- residues 20-134 of Sequence ID No. 2;
- residues 20-134 of Sequence ID No. 2, wherein the arginine residue at position number 39 is replaced by a lysine, and the isoleucine at residue number 92 is replaced by a valine;
- residues 21-132 of Sequence ID No. 4; or
- residues 26-113 of Sequence ID No. 12.

As employed herein, the term "substantially the same amino acid sequence" refers to amino acid sequences having at least about 80% identity with respect to the reference amino acid sequence, and will retain comparable functional and biological properties characteristic of the protein encoded by the reference amino acid. Preferably, proteins having "substantially the same amino acid sequence" will have at least about 90% amino acid identity with respect to the reference amino acid sequence; with greater than about 95% amino acid sequence identity being especially preferred.

The above-described soluble proteins can be employed for a variety of therapeutic uses, e.g., to block receptors of the invention from affecting processes which the receptors would otherwise mediate. The presence of the soluble proteins of the invention will compete with functional ligand for the receptor, preventing the formation of a functional receptor-ligand complex, thereby blocking the normal regulatory action of the complex.

In accordance with yet another embodiment of the present invention, there are provided antibodies generated against the above-described soluble proteins and receptor proteins. Such antibodies can be employed for diagnostic applications, therapeutic applications, and the like. Preferably, for therapeutic applications, the antibodies employed will be monoclonal antibodies.

The above-described antibodies can be prepared employing standard techniques, as are well known to those of skill in the art, using the invention receptor proteins as antigens for antibody production.

In accordance with still another embodiment of the present invention, there are provided methods for modulating the transcription trans-activation of

receptor(s) of the invention by contacting said receptor(s) with a modulating, effective amount of the above-described antibodies.

5 The soluble proteins of the invention, and the antibodies of the invention, can be administered to a subject employing standard methods, such as, for example, by intraperitoneal, intramuscular, intravenous, or subcutaneous injection, implant or transdermal modes of
10 administration, and the like. In addition, methods such as transfection with viral or retroviral vectors encoding the invention compositions. One of skill in the art can readily determine dose forms, treatment regimens, etc, depending on the mode of administration employed.

15 In accordance with a further embodiment of the present invention, there are provided DNA sequences which encode the above-described soluble proteins and receptor proteins. Optionally, such DNA sequences, or fragments
20 thereof, can be labeled with a readily detectable substituent (to be used, for example, as a hybridization probe).

25 The above-described receptor(s) can be encoded by numerous DNA sequences, e.g., a DNA sequence having a contiguous nucleotide sequence substantially the same as:

 nucleotides 128 - 1609 of Sequence ID No. 1
(which encodes a mouse activin receptor);

30 variations of nucleotides 128 - 1609 of Sequence ID No. 1, wherein the codon for residue number 39 of the encoded amino acid codes for lysine, the codon for residue number 92 of the encoded amino acid codes for valine, and the codon for residue number 288 of the encoded amino acid encodes glutamine (which encodes a
35 human activin receptor);

 nucleotides 528 - 1997 of Sequence ID No. 3
(which encodes a Xenopus activin receptor);

variations of any of the above sequences which
encode the same amino acid sequences, but employ
5 different codons for some of the amino acids.

As employed herein, the term "substantially the same as" refers to DNA having at least about 70% homology with respect to the nucleotide sequence of the DNA fragment with which subject DNA is being compared. Preferably, DNA "substantially the same as" a comparative DNA will be at least about 80% homologous to the comparative nucleotide sequence; with greater than about 90% homology being especially preferred.

Another DNA which encodes a receptor of the invention is one having a contiguous nucleotide sequence substantially the same as:

nucleotides 71 - 1609 of Sequence ID No. 1 (which
encodes a precursor-form of a mouse activin receptor);
variations of nucleotides 71 - 1609 of Sequence
ID No. 1, wherein the codon for residue number 39 of
the encoded amino acid codes for lysine, the codon for
residue number 92 of the encoded amino acid codes for
valine, and the codon for residue number 288 of the
encoded amino acid encodes glutamine (which encodes a
precursor-form of a human activin receptor);

30 nucleotides 468 - 1997 of Sequence ID No. 3
 (which encodes a precursor form of a Xenopus activin
 receptor);

nucleotides 72 - 1550 of Sequence ID No. 11
(which encodes a precursor form of a rat activin
receptor); or

variations of any of the above sequences which
35 encode the same amino acid sequences, but employ
different codons for some of the amino acids.

Yet another DNA which encodes the above-described receptor is one having a contiguous nucleotide sequence substantially the same as set forth in Sequence ID No. 1, Sequence ID No. 1', Sequence ID No. 3, or Sequence ID No.

5 11.

In accordance with a further embodiment of the present invention, the receptor-encoding cDNAs can be employed to probe library(ies) (e.g., cDNA, genomic, and
10 the like) for additional sequences encoding novel receptors of the activin/TGF- β superfamily. Such screening is initially carried out under low-stringency conditions, which comprise a temperature of less than about 42°C, a formamide concentration of less than about 50%, and a
15 moderate to low salt concentration. Presently preferred conditions for such screening comprise a temperature of about 37°C, a formamide concentration of about 20%, and a salt concentration of about 5X standard saline citrate (SSC; 20X SSC contains 3M sodium chloride, 0.3M sodium
20 citrate, pH 7.0). Such conditions will allow the identification of sequences which have a substantial degree of similarity with the probe sequence, without requiring perfect homology for the identification of a stable hybrid. The phrase "substantial similarity" refers to sequences
25 which share at least 50% homology. Preferably, hybridization conditions will be selected which allow the identification of sequences having at least 70% homology with the probe, while discriminating against sequences which have a lower degree of homology with the probe.

30

In accordance with yet another embodiment of the present invention, there is provided a method for the recombinant production of receptor(s) of the invention by expressing the above-described DNA sequences in suitable
35 host cells.

The use of a wide variety of recombinant

organisms has been described for the production of peptides. One of skill in the art can readily determine suitable hosts (and expression conditions) for use in the recombinant production of the peptides of the present invention. Yeast hosts, bacterial hosts, mammalian hosts, and the like can be employed. Regulatory sequences capable of controlling the expression of invention peptides are well known for each of these host systems, as are growth conditions under which expression occurs.

In accordance with a further embodiment of the present invention, there is provided a binding assay employing receptors of the invention, whereby a large number of compounds can be rapidly screened to determine which compounds, if any, are capable of binding to the receptors of the invention. Then, more detailed assays can be carried out with those compounds found to bind, to further determine whether such compounds act as agonists or antagonists of invention receptors.

Another application of the binding assay of the invention is the assay of test samples (e.g., biological fluids) for the presence or absence of members of the activin/TGF- β superfamily of polypeptide growth factors. Thus, for example, serum from a patient displaying symptoms related to pathway(s) mediated by members of the activin/TGF- β superfamily of polypeptide growth factors can be assayed to determine if the observed symptoms are perhaps caused by over- or under-production of such polypeptide growth factor.

The binding assays contemplated by the present invention can be carried out in a variety of ways, as can readily be identified by one of skill in the art. For example, competitive binding assays can be employed, as well as radioimmunoassays, ELISA, ERMA, and the like.

In accordance with a still further embodiment of the present invention, there are provided bioassays for evaluating whether test compounds are capable of acting as agonists or antagonists of receptor(s) of the present invention.

The bioassays of the present invention involve evaluating whether test compounds are capable of acting as either agonists or antagonists for members of the invention superfamily of receptors, or functional modified forms of said receptor protein(s). The bioassay for evaluating whether test compounds are capable of acting as agonists comprises:

- (a) culturing cells containing:
 - DNA which expresses said receptor protein(s) or functional modified forms of said receptor protein(s), and
 - DNA encoding a hormone response element operatively linked to a reporter gene;
- wherein said culturing is carried out in the presence of at least one compound whose ability to induce transcription activation activity of receptor protein is sought to be determined, and thereafter
- (b) monitoring said cells for expression of the product of said reporter gene.

The bioassay for evaluating whether test compounds are capable of acting as antagonists for receptor(s) of the invention, or functional modified forms of said receptor(s), comprises:

- (a) culturing cells containing:
 - DNA which expresses said receptor protein(s), or functional modified forms of said receptor protein(s), and
 - DNA encoding a hormone response element operatively linked to a reporter gene

wherein said culturing is carried out in the presence of:

5 increasing concentrations of at least one compound whose ability to inhibit transcription activation of said receptor protein(s) is sought to be determined, and

10 a fixed concentration of at least one agonist for said receptor protein(s), or functional modified forms of said receptor protein(s); and thereafter

15 (b) monitoring in said cells the level of expression of the product of said reporter gene as a function of the concentration of said compound, thereby indicating the ability of said compound to inhibit activation of transcription.

20 Host cells contemplated for use in the bioassay(s) of the present invention, include CV-1 cells, COS cells, and the like; reporter and expression plasmids employed typically also contain the origin of replication of SV-40; and the reporter and expression plasmids employed also typically contain a selectable marker.

25 The hormone response element employed in the bioassay(s) of the present invention can be selected from, for example, mouse mammary tumor virus long terminal repeat (MTV LTR), mammalian growth hormone promoter, and the reporter gene can be selected from chloramphenicol acetyltransferase (CAT), luciferase, β -galactosidase, and
30 the like.

35 The cells can be monitored for the level of expression of the reporter gene in a variety of ways, such as, for example, by photometric means [e.g., by colorimetry (with a colored reporter product such as β -galactosidase), by fluorescence (with a reporter product such as luciferase), etc], by enzyme activity, and the like.

Compounds contemplated for screening in accordance with the invention bioassays include activin- or TGF- β -like compounds, as well as compounds which bear no particular structural or biological relatedness to activin or TGF- β .

As employed herein, the phrase "activin- or TGF- β -like compounds" includes substances which have a substantial degree of homology (at least 20% homology) with the amino acid sequences of naturally occurring mammalian inhibin alpha and β_A or β_B chains (either singly or in any combination) as well as alleles, fragments, homologs or derivatives thereof which have substantially the same qualitative biological activity as mammalian inhibin, activin, or TGF- β . Examples of activin- or TGF- β -like compounds include activin A (a homodimer of two inhibin β_A subunits), activin B (a homodimer of two inhibin β_B subunits), activin AB (a heterodimer composed of one inhibin β_A subunit and one inhibin β_B subunit), inhibin A (composed of the inhibin α subunit and an inhibin β_A subunit), inhibin B (composed of the inhibin α subunit and an inhibin β_B subunit), TGF- β_1 (a homodimer of two TGF- β_1 subunits), TGF- β_2 (a homodimer of two TGF- β_2 subunits), TGF- β_3 (a homodimer of two TGF- β_3 subunits), TGF- β_4 (a homodimer of two TGF- β_4 subunits), TGF- β_5 (a homodimer of two TGF- β_5 subunits), TGF- $\beta_{1.2}$ (a heterodimer of one TGF- β_1 subunit and one TGF- β_2 subunit), and the like.

Examples of compounds which bear no particular structural or biological relatedness to activin or TGF- β , but which are contemplated for screening in accordance with the bioassays of the present invention, include any compound that is capable of either blocking the action of the invention receptor peptides, or promoting the action of the invention receptor peptides, such as, for example, alkaloids and other heterocyclic organic compounds, and the like.

The method employed for cloning the receptor(s) of the present invention involves expressing, in mammalian cells, a cDNA library of any cell type thought to respond to members of the activin/TGF- β superfamily of polypeptide growth factors (e.g., pituitary cells, placental cells, fibroblast cells, and the like). Then, the ability of the resulting mammalian cells to bind a labeled receptor ligand (i.e., a labeled member of the activin/TGF- β superfamily of polypeptide growth factors) is determined. Finally, the desired cDNA insert(s) are recovered, based on the ability of that cDNA, when expressed in mammalian cells, to induce (or enhance) the binding of labeled receptor ligand to said cell.

In addition to the above-described applications of the receptor proteins and DNA sequences of the present invention, the receptor or receptor-encoding compositions of the invention can be used in a variety of ways. For example, since activin is involved in many biological processes, the activin receptor (or antibodies thereto) can be applied to the modulation of such biological processes. For example, the stimulation of FSH release by activin can either be enhanced (for example, by supplying the subject with increased amounts of the activin receptor, relative to the amount of endogenous receptor, e.g., by transfecting the subject with a tissue specific activin-encoding construct), or depressed (e.g., by administration to a subject of antibodies to the activin receptor, thereby preventing formation of activin-receptor complex, which would then act to stimulate the release of FSH). Thus, the compositions of the present invention can be applied to the control of fertility in humans, domesticated animals, and animals of commercial interest.

As another example, the effect of activin on mitosis of red and white blood cells can be modulated, for example, by administering to a subject (employing suitable

means of administration) a modulating, effective amount of activin receptor (which would enhance the ability of activin present in the cell to modulate mitosis). Alternatively, one could administer to a subject an
5 antibody to the activin receptor (or a portion thereof), which would reduce the effect of activin by blocking the normal interaction between activin and activin receptor.

As additional examples of the wide utility of the
10 invention compositions, receptors and/or antibodies of the invention can be used in such areas as the diagnosis and/or treatment of activin-dependent tumors, enhancing the survival of brain neurons, inducing abortion in livestock and other domesticated animals, inducing twinning in
15 livestock and other domesticated animals, and so on.

As still further examples of the wide utility of the invention compositions, agonists identified for TGF- β specific receptors can be used to stimulate wound healing,
20 to suppress the growth of TGF- β -sensitive tumors, to suppress immune response (and thereby prevent rejection of transplanted organs), and the like. Antagonists or the soluble, ligand-binding domain derived from TGF- β receptors can be used to block endogenous TGF- β , thereby promoting
25 liver regeneration and stimulating some immune responses.

It can be readily seen, therefore, that the invention compositions have utility in a wide variety of diagnostic, clinical, veterinary and research applications.

30

The invention will now be described in greater detail by reference to the following non-limiting examples.

EXAMPLES

Recombinant human (rh) activin A, rh activin B, and rh inhibin A were generously provided by Genentech, Inc. Porcine TGF- β 1 was obtained from R+D Systems.

Double-stranded DNA was sequenced by the dideoxy chain termination method using the Sequenase reagents from US Biochemicals. Comparison of DNA sequences to databases was performed using the FASTA program [Pearson and Lipman, Proc. Natl. Acad. Sci. USA 85: 2444-2448 (1988)].

EXAMPLE I

Construction and Subdivision of AtT20 cDNA Library

Polyadenylated RNA was prepared from AtT20 cells using the Fast Track reagents from InVitrogen. cDNA was commercially synthesized and ligated into the plasmid vector pcDNA1 using non-palindromic BstXI linkers, yielding a library of approximately 5×10^6 primary recombinants. The unamplified cDNA library was plated at 1000 clones per 100 mm plate, then scraped off the plates, frozen in glycerol and stored at -70° .

Activin suppresses adrenocorticotrophic hormone (ACTH) secretion by both primary anterior pituitary cell cultures [Vale et al., Nature 321: 776-779 (1986)] and AtT20 mouse corticotropic cells. Because AtT20 cells possess activin receptors indistinguishable from those on other cell types (based on binding affinity measurements with activin A), these cells were chosen to be the source of cDNA for transfection. A cDNA library of approximately 5×10^6 independent clones from AtT20 cells was constructed in the mammalian expression vector, pcDNA1, and screened using an expression cloning approach [Gearing et al., EMBO J. 8, 3667-3676 (1989)] based on the ability to detect activin binding to single transfected cells. The library was

divided into pools of 1000 clones, DNA was prepared from each pool of clones and transiently transfected into COS cells, and the cells screened for the capacity to bind iodinated activin A. Binding was assessed by performing the transfections and binding reactions directly on chambered microscope slides, then dipping the slides in photographic emulsion and analyzing them under a microscope. Cells which had been transfected with an activin receptor cDNA, and consequently bound radioactive activin, were covered with silver grains. DNA from pools of clones were analyzed either singly or in groups of three. Of 300 pools (approximately 300,000 clones) assayed in this manner, one group of three generated two positive cells when transfected into COS cells. The positive pool (#64) was identified by transfecting and analyzing DNA from each pool of 1000 singly, and then was further fractionated until a single clone (pmActR1) was purified which generated >10⁴ positive cells after transfection (see Table 1).

Table 1
Purification of the activin receptor clone from
the AtT20 library

	<u>Pool</u>	<u>Clones/pool</u>	<u>Positive cells/slide</u>
25	62,63,64	3x1000	2
	64	1000	1-3
	64-51	400	4-10
	64-51-R10;64-51-C13	20	25-40
	pmActR1	1	>10 ⁴

The total number of transfected cells capable of binding ¹²⁵I activin A in a field of 2x10⁵ COS cells was counted for pools of clones at each stage of the purification process.

pmActR1 contained a 1.7 kb insert, coding for a protein of 342 amino acids (Figure 3); however, it was

incomplete on the 3' end, thus the last 17 amino acids were encoded by vector sequences. In order to obtain the entire sequence, the AtT20 library was rescreened by hybridization with the 1.6 kb SacI-PstI fragment (Figure 3). Screening
5 6×10^5 colonies yielded one additional positive clone (pmActR2) which had a 2.6 kb insert and contained the entire coding sequence for the mouse activin receptor (Figure 3). The nucleic acid sequence and the deduced amino acid sequence of the insert in pmActR2 are set forth
10 in Sequence ID No. 1.

EXAMPLE II

COS Cell Transfection

15 Aliquots of the frozen pools of clones from Example I were grown overnight in 3 ml cultures of terrific broth, and mini-prep DNA prepared from 1.5 ml using the alkaline lysis method [Maniatis et al. Molecular Cloning (Cold Spring Harbor Laboratory (1982))]. 1/10 of the DNA
20 from a mini-prep (10 Ml of 100 Ml) was used for each transfection.

2x10⁵ COS cells were plated on chambered microscope slides (1 chamber - Nunc) that had been coated
25 with 20 µg/ml poly-D-lysine and allowed to attach for at least 3 hours. Cells were subjected to DEAE-Dextran mediated transfection as follows. 1.5 ml of serum-free Dulbecco's Modified Eagle's medium (DME) containing 100 mM chloroquine was added to the cells. DNA was precipitated
30 in 200 ml DME/chloroquine containing 500 mg/ml DEAE-Dextran, then added to the cells. The cells were incubated at 37° for 4 hours, then the media was removed and the cells were treated with 10% DMSO in HEPES buffered saline for 2 minutes. Fresh media was added and the cells assayed
35 3 days later. For transfections with the purified clone, 2.5x10⁶ cells were transfected in 100 mm dishes with 5 µg purified DNA. The total transfection volume was 10 ml, and

[illegible][illegible][illegible][illegible][illegible][illegible][illegible]

clone, which when transfected yielded $>10^4$ positive cells/ 2×10^5 cells.

EXAMPLE V

5

Radioreceptor Assay

10⁵ COS cells transfected with either pmActR1 or pmActR2, or 10⁶ untransfected COS cells, were plated in 6 well dishes and allowed to grow overnight. The cells were
10 washed 2X with HDB, 0.1% BSA, and incubated at 22° for 90 minutes in 0.5 ml HDB, 0.1% BSA containing 100,000 cpm (approximately 1 ng, 75 pM) ¹²⁵I activin A (5 µg activin A was iodinated by chloramine T oxidation to a specific activity of 50-90 µCi/µg; iodinated activin A was purified
15 on a 0.7x20 cm G-25 column) and varying amounts of unlabeled competitor hormone. Following binding, the cells were washed 3X with cold HDB, solubilized in 0.5 ml 0.5 N NaOH, removed from the dish and radioactivity was measured in a gamma counter. Data presented in Figure 5 are
20 expressed as % specific binding, where 100% specific binding is the difference between binding in the absence of competitor and binding in the presence of a 100 fold molar excess of unlabeled activin A. Binding parameters were determined using the program LIGAND [Munson P.J. and
25 Rodbard, D., Anal. Biochem. 107:220-259 (1980)].

EXAMPLE VI

Chemical Cross-linking

30 2x10⁶ COS cells, or 5x10⁶ AtT20 cells, were washed 2x with HDB, scraped off the dish, incubated for 90 minutes at 22° under constant rotation in 0.5 ml HDB containing 7x10⁵ cpm (approximately 500 pM) ¹²⁵I activin A with or without 500 ng (37 nM) unlabeled activin A. Cells were
35 diluted with 1 ml HDB, pelleted by centrifugation and resuspended in 0.5 ml HDB. Disuccinimidyl suberate (DSS; freshly dissolved in DMSO) was added to 500 µM, and the

cells incubated at 0° for 30 minutes. The cross-linking was terminated by addition of 1 ml 50 mM Tris-HCl pH 7.5, 100 mM NaCl, then the cells were pelleted by centrifugation, resuspended in 100 µl 50 mM Tris-HCl pH 7.5, 1% Triton X-100 and incubated at 0° for 60 minutes. The samples were centrifuged 5 minutes at 13,000xg, and the Triton-soluble supernatants analyzed by SDS-PAGE using 8.5% polyacrylamide gels. The gels were dried and subjected to autoradiography for 4-14 days.

10

EXAMPLE VII
RNA Blot Analysis

Total RNA was purified from tissue culture cells and tissues using LiCl precipitation. 20 µg total RNA was run on 1.2% agarose, 2.2M formaldehyde gels, blotted onto nylon membranes (Hybond - NEN), and hybridized with a 0.6 kb KpnI fragment (see Figure 3) which had been labeled with ³²P by random priming using reagents from US Biochemicals. Hybridization was performed at 42° in 50% formamide, and the filters were washed at 65° in 0.2X SSC.

20

EXAMPLE VIII
Sequence Analysis

Full length mouse activin receptor clone encodes a protein of 513 amino acids, with a 5' untranslated region of 70 bp and a 3' untranslated region of 951 bp. pmActR2 does not contain a poly A tail, although it does have a potential poladenylylation site at bp 2251. The insert in clone pmActR1 had an additional 551 bp of 5' untranslated sequence, was identical in the overlapping range, and stopped at the 3' end at base 1132 of pmActR2. The first methionine codon (ATG), at bp 71, in pmActR2 is in a favorable context for translation initiation [Kozak, M., Nucl. Acids Res. 15:8125-8148 (1987)], and is preceded by an in-frame stop codon. pmActR1 contains 3 additional ATGs

25

30

35

006727 13547460

in the 5' untranslated region; however, none of these is in an appropriate context for initiation, and all are followed by in-frame stop codons. While this unusually long 5' leader sequence may have functional significance, it is clearly not necessary for proper expression, because pmActR2, which lacks most of that sequence, can be functionally expressed in COS cells (see below).

Hydropathy analysis using the method of Kyte and Doolittle [J. Mol. Biol. 157:105-132 (1982)] revealed two hydrophobic regions: a 10 amino acid stretch at the amino terminus assumed to be a single peptide, and a single putative 26 residue membrane-spanning region between amino acids 119-142 (see Figure 1 and Sequence ID No. 2). The signal peptide contains the conserved n-, h- and c- domains common to signal sequences; the site of cleavage of the signal peptide, before Ala¹, is predicted based on rules described by von Heijne [Biochim. Biophys. Act. 947:307-333 (1988)]. As is common for the cytoplasmic side of membrane-spanning domains, the predicted transmembrane region is closely followed by two basic amino acids. The mature mouse activin receptor is thus predicted to be a 494 amino acid type I membrane protein of Mr 54 kDa, with a 116 amino acid N-terminal extracellular ligand binding domain, and a 346 amino acid intracellular signalling domain.

Comparision of the activin receptor sequence to the sequence databases revealed structural similarity in the intracellular domain to a number of receptor and non-receptor kinases. Analysis of the sequences of all kinases has led to the identification of a 300 amino acid kinase domain characterized by 12 subdomains containing a number of highly conserved amino acids [Hanks, S.K. and Quinn, A.M., Meth. Enzymol. 200:38-62 (1991) and Hanks et al., Science 241:42-52 (1988)]; the activin receptor sequence has all of these conserved subdomains in the proper order (Figure 4). A conserved Gly in subdomain I is replaced by

5

The sequences in two of these subdomains (VIB and VIII) can be used to predict tyrosine vs. serine/threonine substrate specificity [Hanks et al., (1988) supra]. The sequence of the mouse activin receptor in both of these
10 subdomains is characteristic of serine kinases.

Table 2
Kinase Domain Predictive Sequences

<u>Subdomain</u>	<u>VIB</u>	<u>SEQ ID NO.</u>	<u>VIII</u>	<u>SEQ ID NO.</u>
serine kinase consensus	DLKPEN	5	G(T/S)XX(Y/F)X	6
activin receptor	DIKSKN	7	GTRRYM	8
tyrosine kinase consensus	DLAARN	9	XP(I/V)(K/R)W(T/M)	10

Therefore, the activin receptor is expected to have serine/threonine specificity. Furthermore, the activin receptor does not have a tyrosine residue in the standard autophosphorylation region between subdomains VII and VIII, indicating that it is not a standard tyrosine kinase. The receptor could potentially autophosphorylate at Ser³³³ or Thr³³⁷. One interesting additional possibility is that the activin receptor kinase may have specificity for serine, threonine and tyrosine residues. Several kinases with these properties have recently been described [see, for example, Howell et al., Mol. Cell. Biol. 11:568-572 (1991), Stern et al., Mol. Cell. Biol. 11:987-1001 (1991) and Featherston, C. and Russell, P., Nature 349:808-811 (1991)].

Phylogenetic analysis of the activin receptor compared to 161 other kinase sequences revealed that the activin receptor and the C.elegans protein, daf-1 [Georgi et al., Cell 61:635-645 (1990)] may constitute a separate subfamily of kinases (see Figure 6). daf-1 is a putative transmembrane receptor involved in the developmental arrest of a non-feeding larval state and shares 32% identity with the activin receptor (see Figure 6). Like the activin receptor, daf-1 is predicted to be a transmembrane serine/threonine-specific kinase; furthermore, both daf and the activin receptor have short, conserved inserts in the kinase domain sequence between subdomains VIA-VIB and X-XI that are not present in any other kinase (underlined in Figure 4B). This additional similarity lends credence to their belonging to a unique subfamily of kinases. The activin receptor is quite distantly related (18% amino acid sequence identity) to the only other known transmembrane serine/threonine protein kinase, encoded by the ZmPK gene of maize [Walker, J.C. and Zhang, R., Nature 345:743-746 (1990)].

The extracellular domain of the activin receptor did not show similarity to any other sequences in the databases. This ligand binding domain is relatively small in comparison to those found in other growth factor
5 receptors, but like those receptors this domain has a high cysteine content. The pattern of these Cys residues, however, is not like either an immunoglobulin fold or the cysteine rich repeats of the EGF receptor. There are also two potential sites of N-linked glycosylation in the
10 extracellular domain, as well as a number of potential phosphorylation sites for protein kinase C and casein kinase II in the intracellular domain.

EXAMPLE IX

15 Binding Properties of the Cloned Activin Receptor

To verify that the cloned receptor is activin specific, competition binding experiments were performed on COS cells transiently transfected with either pmActR1 or
20 pmActR2. Cells transfected with either construct bound activin A with a single high affinity component ($K_d = 180$ pM; Figure 5), indicating that a functional (structurally complete) intracellular kinase domain is not required for ligand binding. This binding affinity is
25 consistent with that measured on other activin-responsive cell types [see, for example, Campen, C.A. and Vale, W., Biochem. Biophys. Res. Comm. 157:844-849 (1988); Hino et al., J. Biol. Chem. 264:10309-10314 (1989); Sugino et al., J. Biol. Chem. 263: 15249-15252 (1988); and Kondo et al.,
30 Biochem. Biophys. Res. Comm, 161:1267-1272 (1989)]. Untransfected COS cells do not bind activin A. The transfected cultures as a whole expressed approximately 26,000 receptors per cell; however, because only 15% of the cells express the transfected gene (as measured by
35 quantitating transfected cells as a fraction of all cells following dipping in emulsion), each transfected cell expressed an average of 175,000 receptors per cell. The

level of expression per cell varies considerably, though, based on the number of accumulated silver grains. This value is comparable to the expression of other transfected cell surface proteins in COS cells.

5

Binding of iodinated activin A to COS cells transiently transfected with pmActR2 could be competed by activin B with slightly reduced potency compared to activin A; by inhibin A with approximately 10-fold lower potency; and could not be competed by TGF- β 1 (Figure 5B). This affinity and specificity of binding match those observed following binding of activin A to a number of other activin-responsive cell types. Although activin B appears to bind the transfected receptor with lower affinity than activin A, the activin B preparation used in these experiments may have suffered a reduction in potency, based on a comparison of bioactivity with activin A, since the recombinant synthesis of the activin B employed herein had been carried out some time ago [recombinant synthesis of activin B is described by Mason et al., in Mol. Endocrinol. 3: 1352-1358 (1989)]. It is likely that this cDNA encodes a receptor for multiple forms of activin.

The size of the cloned activin receptor was analyzed by affinity cross-linking 125 I activin A to COS cells transfected with pmActR2 using the bifunctional chemical cross-linker, disuccinimidyl suberate (DSS). A major cross-linked band of 84 kDa was observed in transfected, but not in untransfected cells. Subtracting the molecular weight of activin, this represents a protein of 56 kDa, which corresponds well to the molecular weight predicted from the nucleic acid sequence data. Cross-linking 125 I activin A to AtT20 cells yields a major band of 65 kDa, with minor bands of approximately 78 and 84 kDa. The size of the largest band matches that generated by the cloned receptor. The smaller bands could be either separate proteins, different phosphorylated forms of the

same protein, or degradation products of the full length
clone; the sequences DKKRR at amino acid 35 and KKKR at
amino acid 416 could be potential sites of proteolysis.
Alternatively, these bands could come from alternatively
5 spliced products of the same gene.

The 84 and 65 kDa cross-linked bands have also
been observed in other activin-responsive cell types [Hino,
supra; Centrella et al., Mol. Cell. Biol. 11:250-258
10 (1991)], and interpreted to represent the signalling
receptor, although complexes of other sizes have also been
seen as well. The size of the activin receptor is very
similar to a putative TGF- β receptor, to the limited extent
it has been characterized by chemical cross-linking [see
15 Massague et al., Ann. N.Y. Acad. Sci. 593: 59-72 (1990)].

EXAMPLE X

Expression of Activin Receptor mRNA

20 The distribution of activin receptor mRNA was
analyzed by Northern blot. Two mRNA species, of 6.0 and
3.0 kb, were observed in AtT20 cells as well as a number of
mouse tissues, including brain, testis, pancreas, liver and
kidney. The total combined size of the inserts from
25 pmActR1 and pmActR2 is 3.1 kb, which corresponds to the
size of the smaller transcript. Neither the extent of
similarity between the two mRNAs, nor the significance of
having two transcripts is clear. The genes for several
other hormone receptors have been shown to be alternatively
30 spliced to generate both a cell surface receptor and a
soluble binding protein, and it is possible that the
activin receptor is processed in a similar manner.

Interestingly, the relative abundance of the two
35 transcripts varies depending on the source. While AtT20
cells have approximately equal levels of both mRNAs, most
tissues had much greater levels of the 6.0 kb transcript,

with little or no expression of the 3.0 kb transcript. Testis, on the other hand, had a greater amount of the 3.0 kb band. Expression of activin receptor mRNA in brain, liver and testis is in accord with described biological actions of activin in those tissues [Mine et al., Endocrinol. 125:586-591 (1989); Vale et al., Peptide Growth Factors and Their Receptors, Handbook of Experimental Pharmacology, M.A. Sporn and A.B. Roberts, ed., Springer-Verlag (1990), in press].

10

EXAMPLE XI

Identification of a Human Activin Receptor

A human testis library (purchased from Clontech; catalog no. HL1010b) was probed with the full length mouse activin receptor gene (see Sequence ID No. 1) under the following conditions:

Hybridization stringency:

20% formamide, 6X SSC at 42°C;

20

Wash stringency:

2X SSC, 0.1% SDS at 42°C.

A sequence which is highly homologous with the mouse activin receptor was identified (Sequence ID No. 1'). Due to the high degree of homology between this receptor and the mouse activin receptor, this receptor is designated as the human form of the activin receptor from the same subclass as the mouse receptor described above.

30

EXAMPLE XII

Identification of a Xenopus Activin Receptor

A Xenopus stage 17 embryo cDNA library (prepared as described by Kintner and Melton in Development 99: 311-325 (1987) was probed with the full length mouse activin receptor gene (see Sequence ID No. 1) under the following conditions:

Hybridization stringency:

20% formamide, 6X SSC at 42°C;

Wash stringency:

2X SSC, 0.1% SDS at 42°C.

5

A sequence having a substantial degree of homology with respect to the mouse activin receptor was identified (Sequence ID No. 3). The degree of overall amino acid homology (relative to the mouse activin receptor) is only about 69% (with 77% homology in the intracellular domain and 58% homology in the extracellular domain). Due to the moderate degree of homology between this receptor and the mouse activin receptor, this receptor is designated as an activin receptor from a different subclass than the mouse receptor described above.

EXAMPLE XIII

Functional Assays of ActRs in Xenopus embryos

To determine whether xActRIIB can transmit a signal in response to activin, xActRIIB RNA was synthesized in vitro and injected into *Xenopus* embryos at two different concentrations. Injected embryos were allowed to develop to stage 9, at which time animal caps were dissected and treated overnight with different concentrations of activin. The xActRIIB cDNA was cloned into rp64T [see Krieg and Melton in Methods in Enzymology, Abelson and Simon, Eds. (Academic Press, New York, 1987), vol. 155, p. 397] and transcribed in vitro to generate a capped, synthetic xActRIIB RNA [see Melton et al., in Nucleic Acids Res. 12:7035 (1984) and Kintner in Neuron 1:545 (1988)]. Embryos at the two- to four-cell stage were injected with about 20 nl of RNA at concentrations of 0.02 ng/nl, or 0.1 ng/nl, spread between four quadrants of the animal pole. At stage 9, animal caps were removed from RNA-injected embryos and incubated in 0.5x modified mammalian Ringer's (MMR), 0.1% bovine serum albumin (BSA) with different

concentrations of purified, porcine activin A (six caps per incubation). After 20 hours in culture, total RNA was prepared.

5 The response of the caps to activin was assessed by quantifying muscle-specific actin RNA with a ribonuclease protection assay as per Blackwell and Weintraub, Science 250:1104 (1990). Embryos injected with 0.4 and 2.0 ng of xActRIIB RNA were approximately 10- and
10 100-fold more sensitive, respectively, to activin than control embryos. The low amount of muscle actin found in animal caps in the absence of added activin A is probably a consequence of contamination of the animal cap with a small amount of marginal zone tissue.

15 The amount of muscle actin decreased with increasing concentration of activin in the embryos injected with 2 ng of xActRIIB RNA. This is consistent with the observation that isolated animal cap cells uniformly
20 exposed to different concentrations of activin only form muscle cells in response to a narrow range of activin concentrations [see Blackmann and Kadesch in Genes and Development 5:1057 (1990)]. The present results indicate that the concentration of ligand and the amount of receptor
25 are both important in determining the signal transmitted. Thus, the range of activin concentrations that lead to muscle differentiation is lower in animal cap cells from injected embryos, which are expressing more receptor than normal, than from uninjected embryos.

30

EXAMPLE XIV

Analysis of kinase activity of mActRII

35 A fragment of cDNA corresponding to the entire intracellular domain of mActRII (amino acids 143-494) was subcloned into the vector pGEX-2T [see Smith and Johnson in Gene 67:31-40 (1988)], creating a fusion protein between

glutathione S-transferase (GST) and the putative kinase domain of the receptor. This plasmid was introduced into bacteria and the expressed fusion protein was purified using glutathione affinity chromatography as described by Smith and Johnson. Approximately 100-200 ng of fusion protein, or of purified GST, were incubated with 25 μ Ci [γ - 32 P] ATP in a buffer containing 50 mM Tris, 10 mM MgCl₂ for 30 minutes at 37°C. The products were analyzed by SDS-PAGE and autoradiography. The fusion protein, but not the GST alone, became phosphorylated, indicating that the kinase domain of the fusion protein was functional. Phosphoamino acid analysis, performed according to Cooper et al. [Meth. Enzym. 99:387 (1983)], indicated that the predominant amino acid residue that became phosphorylated was threonine.

EXAMPLE XV

Identification of a Rat Activin Receptor

Degenerate primers deduced from the conserved serine/threonine kinase domains of activin/TGF β type II receptors were used to perform reverse-transcription polymerase chain reaction (RT-PCR) on a rat cDNA library derived from adult rat pituitary or brain. A mixture of oligo(dT)-primed cDNAs from 5 μ g of total RNA were used as templates for PCR. The degenerate primers used were:

H1: 5'-CGGGATCCGTNGCNGTNAARATHTTYCC-3' (SEQ ID NO:13)
(a sense primer corresponding to amino acid sequence 216-221 of SEQ ID NO:1 in kinase subdomain II); and

H3: 5'-CGGGATCCYTCNGGNGCCATRTANCKYCTNGTNCC-3' (SEQ ID NO:14) (an antisense primer corresponding to amino acid sequence 361-369 of SEQ ID NO:1 in the kinase subdomain VIII).

The primers have BamHI sites at the 5' termini to facilitate the subcloning of the resulting PCR products. The PCR reaction included an initial denaturation step at

94°C for 5 min, 35 cycles of 94°C for 1 min, 46°C for 2 min, and 72°C for 3 min, and a final incubation for 10 min at 72°C. The PCR products were purified and subcloned into the pBluescript vector (Stratagene, La Jolla, CA) and
5 sequenced.

Four fragments having serine/threonine kinase motifs were isolated. Among them, three were previously characterized as ActRI (ALK2), ActRIB (ALK4) and TSRI
10 (ALK1). A full length cDNA of a fourth novel clone from an adult rat brain cDNA library was isolated, and tentatively named ALK7 (activin receptor-like kinase 7). The nucleotide and amino acid sequences for ALK7 are set forth in SEQ ID NOs:11 and 12.

15 The kinase domain of ALK7 shows highest sequence similarity to that of ActRIB and TGF β RI (82.5% identities with them), and the entire amino acid sequence shows 64.0% identity to that of TGF β RI, and 62.1% identity to that of
20 ActRIB. Furthermore, ALK7 has a "GS domain" almost identical to TGF β RI and ActRIB, and contains cysteine residues in the extracellular ligand binding domain conserved among the receptor serine kinase superfamily. This indicates that ALK7 may function as a type I receptor
25 for the TGF- β superfamily.

30 RNase protection assays using RNAs isolated from various rat brain, kidney, stomach, spleen, heart, skin, skeletal muscle, ovary and testis were conducted to determine the expression patterns of the ALK7 gene. Although ALK7 mRNA is not expressed at a high level in adult tissues, it is clearly detectable in brain and to a lesser extent in kidney and ovary.

35 Functional characterization of ALK7 or an ALK7 mutant ALK7(T194D)) was performed in the mink lung cell-line "R1B", Chinese Hamster Ovary cell-line (CHO), and human

20 While the invention has been described in detail
with reference to certain preferred embodiments thereof, it
will be understood that modifications and variations are
within the spirit and scope of that which is described and
claimed.

25

SUMMARY OF SEQUENCES

Sequence ID No. 1 is the nucleic acid sequence
(and the deduced amino acid sequence) of a cDNA encoding a
5 mouse-derived activin receptor of the present invention.

Sequence ID No. 1' is a nucleic acid sequence
encoding a human-derived activin receptor of the present
invention. Sequence ID No. 1' is substantially the same as
10 Sequence ID No. 1, except that the codon for amino acid
residue number 39 encodes lysine (i.e., nucleotides 185-187
are AAA or AAG), the codon for amino acid residue 92
encodes valine (i.e., nucleotides 344-346 are GTN, wherein
N is A, C, G or T), and the codon for amino acid residue
15 number 288 encodes glutamine (i.e., nucleotides 932-934 are
CAA or CAG).

Sequence ID No. 2 is the deduced amino acid
sequence of a mouse-derived activin receptor of the present
20 invention.

Sequence ID No. 2' is an amino acid sequence for
a human-derived activin receptor of the present invention.
Sequence ID No. 2' is substantially the same as Sequence ID
25 No. 2, except that amino acid residue number 39 is lysine,
amino acid residue 92 is valine, and amino acid residue
number 288 is glutamine.

Sequence ID No. 3 is the nucleic acid sequence
30 (and the deduced amino acid sequence) of a cDNA encoding a
Xenopus-derived activin receptor of the present invention.

Sequence ID No. 4 is the deduced amino acid
sequence of a Xenopus-derived activin receptor of the
35 present invention.

Sequence ID No. 5 is the amino acid sequence of the VIB subdomain of the serine kinase consensus sequence.

Sequence ID No. 6 is the amino acid sequence of
5 the VIII subdomain of the serine kinase consensus sequence.

Sequence ID No. 7 is the amino acid sequence of the VIB subdomain of the invention activin receptor.

Sequence ID No. 8 is the amino acid sequence of the VIII subdomain of the invention activin receptor.

Sequence ID No. 9 is the amino acid sequence of
the VIB subdomain of the tyrosine kinase consensus
15 sequence.

Sequence ID No. 10 is the amino acid sequence of the VIII subdomain of the tyrosine kinase consensus sequence.

Sequence ID No. 11 is the nucleic acid sequence (and the deduced amino acid sequence) of a cDNA encoding rat-derived activin receptor of the present invention.

25 Sequence ID No. 12 is the deduced amino acid
sequence of a rat-derived activin receptor of the present
invention.

Sequence ID No. 13 is the H1 degenerate primer
30 employed in Example XV.

Sequence ID No. 14 is the H3 degenerate primer employed in Example XV.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Mathews, Lawrence S.
Vale, Wylie W.
Tsuchida, Kunihiro
- (ii) TITLE OF INVENTION: CLONING AND RECOMBINANT PRODUCTION OF
RECEPTOR(S) OF THE ACTIVIN/TGF-BETA SUPERFAMILY
- (iii) NUMBER OF SEQUENCES: 14
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
 - (B) STREET: 444 South Flower Street, Suite 2000
 - (C) CITY: Los Angeles
 - (D) STATE: CA
 - (E) COUNTRY: USA
 - (F) ZIP: 90071
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/300,584
 - (B) FILING DATE: 02-SEP-1994
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/880,220
 - (B) FILING DATE: 08-MAY-1992
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/773,229
 - (B) FILING DATE: 09-OCT-1991
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/698,709
 - (B) FILING DATE: 10-MAY-1991
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Reiter, Stephen E.
 - (B) REGISTRATION NUMBER: 31,192
 - (C) REFERENCE/DOCKET NUMBER: P41 9927
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 619-546-4737
 - (B) TELEFAX: 619-546-9392

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2563 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA

(A) NAME/KEY: CDS
(B) LOCATION: 71..1609

CTCCGAGGAA	GACCCAGGGA	ACTGGATATC	TAGCGAGAAC	TTCCTACGGC	TTCTCCGGCG		60
CCTCGGGAAA	ATG GGA GCT GCT GCA AAG TTG GCG TTC GCC GTC TTT CTT					109	
	Met Gly Ala Ala Ala Lys Leu Ala Phe Ala Val Phe Leu	1	5	10			
ATC TCT TGC TCT TCA GGT GCT ATA CTT GGC AGA TCA GAA ACT CAG GAG						157	
Ile Ser Cys Ser Ser Gly Ala Ile Leu Gly Arg Ser Glu Thr Gln Glu	15	20	25				
TGT CTT TTC TTT AAT GCT AAT TGG GAA AGA GAC AGA ACC AAC CAG ACT						205	
Cys Leu Phe Phe Asn Ala Asn Trp Glu Arg Asp Arg Thr Asn Gln Thr	30	35	40	45			
GGT GTT GAA CCT TGC TAT GGT GAT AAA GAT AAA CGG CGA CAT TGT TTT						253	
Gly Val Glu Pro Cys Tyr Gly Asp Lys Asp Lys Arg Arg His Cys Phe	50	55	60				
GCT ACC TGG AAG AAT ATT TCT GGT TCC ATT GAA ATA GTG AAG CAA GGT						301	
Ala Thr Trp Lys Asn Ile Ser Gly Ser Ile Glu Ile Val Lys Gln Gly	65	70	75				
TGT TGG CTG GAT GAT ATC AAC TGC TAT GAC AGG ACT GAT TGT ATA GAA						349	
Cys Trp Leu Asp Asp Ile Asn Cys Tyr Asp Arg Thr Asp Cys Ile Glu	80	85	90				
AAA AAA GAC AGC CCT GAA GTG TAC TTT TGT TGC TGT GAG GGC AAT ATG						397	
Lys Lys Asp Ser Pro Glu Val Tyr Phe Cys Cys Cys Glu Gly Asn Met	95	100	105				
TGT AAT GAA AAG TTC TCT TAT TTT CCG GAG ATG GAA GTC ACA CAG CCC						445	
Cys Asn Glu Lys Phe Ser Tyr Phe Pro Glu Met Glu Val Thr Gln Pro	110	115	120	125			
ACT TCA AAT CCT GTT ACA CCG AAG CCA CCC TAT TAC AAC ATT CTG CTG						493	
Thr Ser Asn Pro Val Thr Pro Lys Pro Pro Tyr Tyr Asn Ile Leu Leu	130	135	140				
TAT TCC TTG GTA CCA CTA ATG TTA ATT GCA GGA ATT GTC ATT TGT GCA						541	
Tyr Ser Leu Val Pro Leu Met Leu Ile Ala Gly Ile Val Ile Cys Ala	145	150	155				
TTT TGG GTG TAC AGA CAT CAC AAG ATG GCC TAC CCT CCT GTA CTT GTT						589	
Phe Trp Val Tyr Arg His His Lys Met Ala Tyr Pro Pro Val Leu Val	160	165	170				
CCT ACT CAA GAC CCA GGA CCA CCC CCA CCT TCC CCA TTA CTA GGG TTG						637	
Pro Thr Gln Asp Pro Gly Pro Pro Pro Pro Ser Pro Leu Leu Gly Leu	175	180	185				
AAG CCA TTG CAG CTG TTA GAA GTG AAA GCA AGG GGA AGA TTT GGT TGT						685	
Lys Pro Leu Gln Leu Leu Glu Val Lys Ala Arg Gly Arg Phe Gly Cys	190	195	200	205			
GTC TGG AAA GCC CAG TTG CTC AAT GAA TAT GTG GCT GTC AAA ATA TTT						733	
Val Trp Lys Ala Gln Leu Leu Asn Glu Tyr Val Ala Val Lys Ile Phe	210	215	220				

CCA Pro	ATA Ile	CAG Gln	GAC Asp 225	AAA Lys	CAG Gln	TCC Ser	TGG Trp	CAG Gln 230	AAT Asn	GAA Glu	TAT Tyr	GAA Glu 235	GTC Val	TAT Tyr	AGT Ser	781
CTA Leu	CCT Pro	GGA Gly 240	ATG Met	AAG Lys	CAT His	GAG Glu	AAC Asn 245	ATA Ile	CTA Leu	CAG Gln	TTC Phe	ATT Ile 250	GGT Gly	GCA Ala	GAG Glu	829
AAA Lys	AGA Arg 255	GGC Gly	ACC Thr	AGT Ser	GTG Val	GAT Asp 260	GTG Val	GAC Asp	CTG Leu	TGG Trp	CTA Leu 265	ATC Ile	ACA Thr	GCA Ala	TTT Phe	877
CAT His 270	GAA Glu	AAG Lys	GGC Gly	TCA Ser	CTG Leu 275	TCA Ser	GAC Asp	TTT Phe	CTT Leu	AAG Lys 280	GCT Ala	AAT Asn	GTG Val	GTC Val	TCT Ser 285	925
TGG Trp	AAT Asn	GAA Glu	CTT Leu	TGT Cys 290	CAT His	ATT Ile	GCA Ala	GAA Glu	ACC Thr 295	ATG Met	GCT Ala	AGA Arg	GGA Gly	TTG Leu 300	GCA Ala	973
TAT Tyr	TTA Leu	CAT His 305	GAG Glu	GAT Asp	ATA Ile	CCT Pro	GGC Gly	TTA Leu 310	AAA Lys	GAT Asp	GGC Gly	CAC His 315	AAG Lys	CCT Pro	GCA Ala	1021
ATC Ile	TCT Ser 320	CAC His	AGG Arg	GAC Asp	ATC Ile	AAA Lys	AGT Ser 325	AAA Lys	AAT Asn	GTG Val	CTG Leu	TTG Leu 330	AAA Lys	AAC Asn	AAT Asn	1069
CTG Leu 335	ACA Thr	GCT Ala	TGC Cys	ATT Ile	GCT Ala	GAC Asp 340	TTT Phe	GGG Gly	TTG Leu	GCC Ala	TTA Leu 345	AAG Lys	TTC Phe	GAG Glu	GCT Ala	1117
GGC Gly 350	AAG Lys	TCT Ser	GCA Ala	GGT Gly	GAC Asp 355	ACC Thr	CAT His	GGG Gly	CAG Gln	GTT Val 360	GGT Gly	ACC Thr	CGG Arg	AGG Arg	TAT Tyr 365	1165
ATG Met	GCT Ala	CCA Pro	GAG Glu	GTG Val 370	TTG Leu	GAG Glu	GGT Gly	GCT Ala	ATA Ile 375	AAC Asn	TTC Phe	CAA Gln	AGG Arg 380	GAC Asp	GCA Ala	1213
TTT Phe	CTG Leu	AGG Arg 385	ATA Ile	GAT Asp	ATG Met	TAC Tyr	GCC Ala	ATG Met 390	GGA Gly	TTA Leu	GTC Val	CTA Leu 395	TGG Trp	GAA Glu	TTG Leu	1261
GCT Ala	TCT Ser 400	CGT Arg	TGC Cys	ACT Thr	GCT Ala	GCA Ala	GAT Asp 405	GGA Gly	CCC Pro	GTA Val	GAT Asp	GAG Glu 410	TAC Tyr	ATG Met	TTA Leu	1309
CCA Pro	TTT Phe 415	GAG Glu	GAA Glu	GAA Glu	ATT Ile	GGC Gly 420	CAG Gln	CAT His	CCA Pro	TCT Ser	CTT Leu 425	GAA Glu	GAT Asp	ATG Met	CAG Gln	1357
GAA Glu 430	GTT Val	GTT Val	GTG Val	CAT His	AAA Lys 435	AAA Lys	AAG Lys	AGG Arg	CCT Pro	GTT Val 440	TTA Leu	AGA Arg	GAT Asp	TAT Tyr	TGG Trp 445	1405
CAG Gln	AAA Lys	CAT His	GCA Ala 450	GGA Gly	ATG Met	GCA Ala	ATG Met	CTC Leu	TGT Cys 455	GAA Glu	ACG Thr	ATA Ile	GAA Glu 460	GAA Glu	TGT Cys	1453
TGG Trp	GAT Asp	CAT His 465	GAT Asp	GCA Ala	GAA Glu	GCC Ala	AGG Arg	TTA Leu 470	TCA Ser	GCT Ala	GGA Gly	TGT Cys 475	GTA Val	GGT Gly	GAA Glu	1501
AGA Arg	ATT Ile 480	ACT Thr	CAG Gln	ATG Met	CAA Gln	AGA Arg	CTA Leu 485	ACA Thr	AAT Asn	ATC Ile	ATT Ile 490	ACT Thr	ACA Thr	GAG Glu	GAC Asp	1549

ATT GTA ACA GTG GTC ACA ATG GTG ACA AAT GTT GAC TTT CCT CCC AAA	1597
Ile Val Thr Val Val Thr Met Val Thr Asn Val Asp Phe Pro Pro Lys	
495 500 505	
GAA TCT AGT CTA TGATGGTGGC ACCGTCTGTA CACACTGAGG ACTGGGACTC	1649
Glu Ser Ser Leu	
510	
TGAACTGGAG CTGCTAAGCT AAGGAAAGTG CTTAGTTGAT TTTCTGTGTG AAATGAGTAG	1709
GATGCCTCCA GGACATGTAC GCAAGCAGCC CCTTGTGGAA AGCATGGATC TGGGAGATGG	1769
ATCTGGGAAA CTTACTGCAT CGTCTGCAGC ACAGATATGA AGAGGAGTCT AAGGGAAAAG	1829
CTGCAAACTG TAAAGAACTT CTGAAAATGT ACTCGAAGAA TGTGGCCCTC TCCAAATCAA	1889
GGATCTTTTG GACCTGGCTA ATCAAGTATT TGCAAACTG ACATCAGATT TCTTAATGTC	1949
TGTCAGAAGA CACTAATTCC TTAAATGAAC TACTGCTATT TTTTAAAT GAAAACTTT	2009
TCATTTTCTG TTTTAAAAAG GGTAACTTTT TATTGCATTT GCTGTTGTTT CTATAAATGA	2069
CTATTGTAAAT GCCAACATGA CACAGCTTGT GAATGTGTAG TGTGCTGCTG TTCTGTGTAC	2129
ATAGTCATCA AAGTGGGGTA CAGTAAAGAG GCTTCCAAGC ATTACTTTAA CCTCCCTCAA	2189
CAAGGTATAC CTCAGTTCCA CGGTTGTTAA ATTATAAAAT TGAAAACACT AACAGAATTT	2249
GAATAAATCA GTCCATGTTT TATAACAAGG TTAATTACAA ATTCAGTGTG TTATTTAAGA	2309
AAAAATGGTA AGCTATGCTT AGTGCCAATA GTAAGTGGCT ATTTGTAAAG CAGTGTTTTA	2369
GCTTTTCTTC TACTGGCTTG TAATTTAGGG AAAACAAGTG CTGTCTTTGA AATGGAAAAG	2429
AATATGGTGT CACCCTACCC CCCATACTTA TATCAAGGTC CCAAATATT CTTTTCCATT	2489
TCAAAGACAG CACTTTGAAA ACCCTAAATT ACAAGCCAGT AGAAGAAAAG CTAAACACG	2549
CTTTACAAAT AGCC	2563

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 513 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Gly	Ala	Ala	Ala	Lys	Leu	Ala	Phe	Ala	Val	Phe	Leu	Ile	Ser	Cys
1				5					10					15	
Ser	Ser	Gly	Ala	Ile	Leu	Gly	Arg	Ser	Glu	Thr	Gln	Glu	Cys	Leu	Phe
		20					25						30		
Phe	Asn	Ala	Asn	Trp	Glu	Arg	Asp	Arg	Thr	Asn	Gln	Thr	Gly	Val	Glu
		35				40						45			
Pro	Cys	Tyr	Gly	Asp	Lys	Asp	Lys	Arg	Arg	His	Cys	Phe	Ala	Thr	Trp
	50					55					60				
Lys	Asn	Ile	Ser	Gly	Ser	Ile	Glu	Ile	Val	Lys	Gln	Gly	Cys	Trp	Leu
65					70					75					80

Asp	Asp	Ile	Asn	Cys 85	Tyr	Asp	Arg	Thr	Asp 90	Cys	Ile	Glu	Lys	Lys 95	Asp
Ser	Pro	Glu	Val 100	Tyr	Phe	Cys	Cys	Cys 105	Glu	Gly	Asn	Met	Cys 110	Asn	Glu
Lys	Phe	Ser 115	Tyr	Phe	Pro	Glu	Met 120	Glu	Val	Thr	Gln	Pro 125	Thr	Ser	Asn
Pro	Val 130	Thr	Pro	Lys	Pro	Pro 135	Tyr	Tyr	Asn	Ile	Leu 140	Leu	Tyr	Ser	Leu
Val 145	Pro	Leu	Met	Leu	Ile 150	Ala	Gly	Ile	Val	Ile 155	Cys	Ala	Phe	Trp	Val 160
Tyr	Arg	His	His	Lys 165	Met	Ala	Tyr	Pro	Pro 170	Val	Leu	Val	Pro	Thr 175	Gln
Asp	Pro	Gly	Pro 180	Pro	Pro	Pro	Ser	Pro 185	Leu	Leu	Gly	Leu	Lys 190	Pro	Leu
Gln	Leu	Leu 195	Glu	Val	Lys	Ala	Arg 200	Gly	Arg	Phe	Gly	Cys 205	Val	Trp	Lys
Ala	Gln 210	Leu	Leu	Asn	Glu	Tyr 215	Val	Ala	Val	Lys	Ile 220	Phe	Pro	Ile	Gln
Asp 225	Lys	Gln	Ser	Trp	Gln 230	Asn	Glu	Tyr	Glu	Val 235	Tyr	Ser	Leu	Pro	Gly 240
Met	Lys	His	Glu	Asn 245	Ile	Leu	Gln	Phe	Ile 250	Gly	Ala	Glu	Lys	Arg 255	Gly
Thr	Ser	Val	Asp 260	Val	Asp	Leu	Trp	Leu	Ile 265	Thr	Ala	Phe	His 270	Glu	Lys
Gly	Ser	Leu 275	Ser	Asp	Phe	Leu	Lys 280	Ala	Asn	Val	Val	Ser 285	Trp	Asn	Glu
Leu	Cys 290	His	Ile	Ala	Glu	Thr 295	Met	Ala	Arg	Gly	Leu 300	Ala	Tyr	Leu	His
Glu 305	Asp	Ile	Pro	Gly	Leu 310	Lys	Asp	Gly	His	Lys 315	Pro	Ala	Ile	Ser	His 320
Arg	Asp	Ile	Lys 325	Ser	Lys	Asn	Val	Leu	Leu 330	Lys	Asn	Asn	Leu	Thr 335	Ala
Cys	Ile	Ala	Asp 340	Phe	Gly	Leu	Ala	Leu 345	Lys	Phe	Glu	Ala	Gly 350	Lys	Ser
Ala	Gly	Asp 355	Thr	His	Gly	Gln	Val 360	Gly	Thr	Arg	Arg	Tyr 365	Met	Ala	Pro
Glu	Val 370	Leu	Glu	Gly	Ala	Ile 375	Asn	Phe	Gln	Arg	Asp 380	Ala	Phe	Leu	Arg
Ile 385	Asp	Met	Tyr	Ala	Met 390	Gly	Leu	Val	Leu	Trp 395	Glu	Leu	Ala	Ser	Arg 400
Cys	Thr	Ala	Ala	Asp 405	Gly	Pro	Val	Asp	Glu 410	Tyr	Met	Leu	Pro	Phe 415	Glu
Glu	Glu	Ile	Gly 420	Gln	His	Pro	Ser	Leu 425	Glu	Asp	Met	Gln	Glu 430	Val	Val

Val His Lys Lys Lys Arg Pro Val Leu Arg Asp Tyr Trp Gln Lys His
 435 440 445
 Ala Gly Met Ala Met Leu Cys Glu Thr Ile Glu Glu Cys Trp Asp His
 450 455 460
 Asp Ala Glu Ala Arg Leu Ser Ala Gly Cys Val Gly Glu Arg Ile Thr
 465 470 475 480
 Gln Met Gln Arg Leu Thr Asn Ile Ile Thr Thr Glu Asp Ile Val Thr
 485 490 495
 Val Val Thr Met Val Thr Asn Val Asp Phe Pro Pro Lys Glu Ser Ser
 500 505 510
 Leu

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2335 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vii) IMMEDIATE SOURCE:
 (B) CLONE: XACTR

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 468..1997

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CCGCCCACAC AGTGCACTGA ATAATAGCCG GTGCGGCCCC TCCCCTCTTT CCCTGGCAGT	60
TGTGTATCTG TCACATTGAA GTTTGGGCTC CTGTGAGTCT GAGCCTCCCC CTGTGTCTCA	120
TGTGAAGCTG CTGCTGCAGA AGGTGGAGTC GTTGCATGAG GGTGGGGGGA GTCGCTGCTG	180
TTTGATCTGC CTCTGCTCCC CATTACACT CTCATTTTCAT TCCCACGGAT CCACATTACA	240
ACTCGCCTTT AACCCCTTCC CTGGCGGAGC CCACGCGTCT TTCATCCCTC CTGCCGCGGC	300
CGCTGAGCGA CCAGAGCGCG ACATTGTTGC GGCAGGGGAT TGGGCGACAT TGTTGCGAAT	360
AATCGGAGCT GCTGGGGGGG AACTGATACA ACGTTGCGAC TGTAAGAGAA TTAACCTCGGC	420
CGAATGGGAT TTTATCTGTG TCGGTGAGAG AAGCGGATCC CAGGAGC ATG GGG GCG	476
	Met Gly Ala
	1
TCT GTA GCG CTG ACT TTT CTA CTT CTT CTT GCA ACT TTC CGC GCA GGC	524
Ser Val Ala Leu Thr Phe Leu Leu Leu Leu Ala Thr Phe Arg Ala Gly	
5 10 15	
TCA GGA CAC GAT GAA GTG GAG ACA AGA GAG TGC ATC TAT TAC AAT GCC	572
Ser Gly His Asp Glu Val Glu Thr Arg Glu Cys Ile Tyr Tyr Asn Ala	
20 25 30 35	

AAC Asn	TGG Trp	GAA Glu	CTG Leu	GAG Glu 40	AAG Lys	ACC Thr	AAC Asn	CAA Gln	AGT Ser 45	GGG Gly	GTG Val	GAA Glu	AGC Ser	TGC Cys 50	GAA Glu	620
GGG Gly	GAA Glu	AAG Lys	GAC Asp 55	AAG Lys	CGA Arg	CTC Leu	CAC His	TGT Cys 60	TAC Tyr	GCG Ala	TCT Ser	TGG Trp	AGG Arg 65	AAC Asn	AAT Asn	668
TCG Ser	GGC Gly	TTC Phe 70	ATA Ile	GAG Glu	CTG Leu	GTG Val	AAA Lys 75	AAA Lys	GGA Gly	TGC Cys	TGG Trp	CTG Leu	GAT Asp	GAC Asp	TTC Phe	716
AAC Asn	TGT Cys 85	TAT Tyr	GAC Asp	AGA Arg	CAG Gln	GAA Glu 90	TGT Cys	ATT Ile	GCC Ala	AAG Lys	GAA Glu 95	GAA Glu	AAC Asn	CCC Pro	CAA Gln	764
GTC Val 100	TTT Phe	TTC Phe	TGC Cys	TGC Cys 105	GAG Glu	GGA Gly	AAC Asn	TAC Tyr	TGC Cys 110	AAC Asn	AAG Lys	AAA Lys	TTT Phe	ACT Thr 115		812
CAT His	TTG Leu	CCT Pro	GAA Glu 120	GTC Val	GAA Glu	ACA Thr	TTT Phe	GAT Asp	CCG Pro 125	AAG Lys	CCC Pro	CAG Gln	CCG Pro	TCA Ser 130	GCC Ala	860
TCC Ser	GTA Val	CTG Leu	AAC Asn 135	ATT Ile	CTG Leu	ATC Ile	TAT Tyr	TCC Ser 140	CTG Leu	CTT Leu	CCA Pro	ATT Ile	GTT Val 145	GGT Gly	CTT Leu	908
TCC Ser	ATG Met 150	GCA Ala	ATT Ile	CTC Leu	CTG Leu	GCG Ala	TTC Phe 155	TGG Trp	ATG Met	TAC Tyr	CGT Arg	CAT His 160	CGA Arg	AAG Lys	CCT Pro	956
CCC Pro	TAC Tyr 165	GGG Gly	CAT His	GTA Val	GAG Glu	ATC Ile 170	AAT Asn	GAG Glu	GAC Asp	CCC Pro	GGT Gly 175	CTG Leu	CCC Pro	CCT Pro	CCA Pro	1004
TCT Ser 180	CCT Pro	CTG Leu	GTC Val	GGG Gly 185	CTG Leu	AAG Lys	CCG Pro	CTG Leu	CAG Gln	TTG Leu 190	CTG Leu	GAG Glu	ATA Ile	AAG Lys	GCG Ala 195	1052
CGA Arg	GGC Gly	CGT Arg	TTC Phe 200	GGT Gly	TGC Cys	GTC Val	TGG Trp	AAA Lys	GCT Ala 205	CGT Arg	CTG Leu	CTG Leu	AAT Asn	GAA Glu 210	TAT Tyr	1100
GTC Val	GCA Ala	GTG Val	AAA Lys 215	ATC Ile	TTC Phe	CCC Pro	GTG Val	CAG Gln 220	GAT Asp	AAG Lys	CAG Gln	TCG Ser	TGG Trp 225	CAG Gln	TGT Cys	1148
GAG Glu	AAA Lys 230	GAG Glu	ATC Ile	TTC Phe	ACC Thr	ACG Thr	CCG Pro 235	GGC Gly	ATG Met	AAA Lys	CAT His	GAA Glu 240	AAC Asn	CTA Leu	TTG Leu	1196
GAG Glu	TTC Phe 245	ATT Ile	GCC Ala	GCT Ala	GAG Glu	AAG Lys 250	AGG Arg	GGA Gly	AGC Ser	AAC Asn	CTG Leu 255	GAG Glu	ATG Met	GAG Glu	CTG Leu	1244
TGG Trp 260	CTC Leu	ATC Ile	ACT Thr	GCA Ala 265	TTT Phe	CAT His	GAT Asp	AAG Lys	GGT Gly 270	TCT Ser	CTG Leu	ACG Thr	GAC Asp	TAC Tyr	CTG Leu 275	1292
AAA Lys	GGG Gly	AAC Asn	TTG Leu 280	GTG Val	AGC Ser	TGG Trp	AAT Asn	GAA Glu 285	CTG Leu	TGT Cys	CAC His	ATA Ile	ACA Thr	GAA Glu 290	ACA Thr	1340
ATG Met	GCT Ala	CGT Arg	GGG Gly 295	CTG Leu	GCC Ala	TAC Tyr	TTA Leu 300	CAT His	GAA Glu	GAT Asp	GTG Val	CCC Pro	CGC Arg 305	TGT Cys	AAA Lys	1388

GGT GAA GGG CAC AAA CCT GCA ATC GCT CAC AGA GAT TTT AAA AGT AAG Gly Glu Gly His Lys Pro Ala Ile Ala His Arg Asp Phe Lys Ser Lys	1436
310 315 320	
AAT GTA TTG CTA AGA AAC GAC CTG ACT GCG ATA TTA GCA GAC TTC GGG Asn Val Leu Leu Arg Asn Asp Leu Thr Ala Ile Leu Ala Asp Phe Gly	1484
325 330 335	
CTG GCC GTA CGA TTT GAG CCT GGA AAA CCT CCG GGA GAT ACA CAC GGG Leu Ala Val Arg Phe Glu Pro Gly Lys Pro Pro Gly Asp Thr His Gly	1532
340 345 350 355	
CAG GTT GGC ACC AGG AGG TAT ATG GCT CCT GAG GTT CTA GAG GGA GCA Gln Val Gly Thr Arg Arg Tyr Met Ala Pro Glu Val Leu Glu Gly Ala	1580
360 365 370	
ATT AAC TTT CAG CGA GAT TCC TTT CTC AGG ATA GAT ATG TAT GCC ATG Ile Asn Phe Gln Arg Asp Ser Phe Leu Arg Ile Asp Met Tyr Ala Met	1628
375 380 385	
GGA CTG GTA CTC TGG GAA ATA GTA TCC CGA TGT ACA GCA GCA GAT GGG Gly Leu Val Leu Trp Glu Ile Val Ser Arg Cys Thr Ala Ala Asp Gly	1676
390 395 400	
CCA GTA GAT GAG TAT CTG CTC CCA TTC GAA GAA GAG ATT GGG CAA CAT Pro Val Asp Glu Tyr Leu Leu Pro Phe Glu Glu Glu Ile Gly Gln His	1724
405 410 415	
CCT TCC CTA GAG GAT CTG CAA GAA GTT GTC GTT CAC AAG AAG ATA CGC Pro Ser Leu Glu Asp Leu Gln Glu Val Val Val His Lys Lys Ile Arg	1772
420 425 430 435	
CCT GTA TTC AAA GAC CAC TGG CTG AAA CAC CCT GGT CTG GCC CAA CTG Pro Val Phe Lys Asp His Trp Leu Lys His Pro Gly Leu Ala Gln Leu	1820
440 445 450	
TGC GTC ACC ATT GAA GAA TGC TGG GAC CAT GAT GCG GAA GCA CGG CTT Cys Val Thr Ile Glu Glu Cys Trp Asp His Asp Ala Glu Ala Arg Leu	1868
455 460 465	
TCG GCA GGC TGC GTA GAG GAG CGT ATT TCC CAA ATC CGT AAA TCA GTG Ser Ala Gly Cys Val Glu Glu Arg Ile Ser Gln Ile Arg Lys Ser Val	1916
470 475 480	
AAC GGC ACT ACC TCG GAC TGC CTT GTA TCC ATT GTT ACA TCT GTC ACC Asn Gly Thr Thr Ser Asp Cys Leu Val Ser Ile Val Thr Ser Val Thr	1964
485 490 495	
AAT GTG GAC TTG CCG CCC AAA GAG TCC AGT ATC TGAGGTTTCT TTGGTCTTTC Asn Val Asp Leu Pro Pro Lys Glu Ser Ser Ile	2017
500 505 510	
CAGACTCAGT GACTTTTAAA AAAAAAATC ACGAATGCAG CTGCTATTTT ATCTTGACTT	2077
TTTAATATTT TTTTCTTGG ATTTACTTG GATCGGATCA ATTTACCAGC ACGTCATTCG	2137
AAAGTATTAA AAAAAAAAAA CAAAACAAAA AAGCAAAAAC AGACATCTCA GCAAGCATTC	2197
AGGTGCCGAC TTATGAATGC CAATAGGTGC AGGAACTTCA GAACCTCAAC AAACCTATTT	2257
CTAGAGAATG TTCTCCTGGT TTCCTTTATC TCAGAAGAGG ACCCATAGGA AAACACCTAA	2317
GTCAAGCAAA TGCTGCAG	2335

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 510 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Met Gly Ala Ser Val Ala Leu Thr Phe Leu Leu Leu Ala Thr Phe
 1           5           10           15
Arg Ala Gly Ser Gly His Asp Glu Val Glu Thr Arg Glu Cys Ile Tyr
          20           25           30
Tyr Asn Ala Asn Trp Glu Leu Glu Lys Thr Asn Gln Ser Gly Val Glu
          35           40           45
Ser Cys Glu Gly Glu Lys Asp Lys Arg Leu His Cys Tyr Ala Ser Trp
          50           55           60
Arg Asn Asn Ser Gly Phe Ile Glu Leu Val Lys Lys Gly Cys Trp Leu
          65           70           75           80
Asp Asp Phe Asn Cys Tyr Asp Arg Gln Glu Cys Ile Ala Lys Glu Glu
          85           90           95
Asn Pro Gln Val Phe Phe Cys Cys Cys Glu Gly Asn Tyr Cys Asn Lys
          100          105          110
Lys Phe Thr His Leu Pro Glu Val Glu Thr Phe Asp Pro Lys Pro Gln
          115          120          125
Pro Ser Ala Ser Val Leu Asn Ile Leu Ile Tyr Ser Leu Leu Pro Ile
          130          135          140
Val Gly Leu Ser Met Ala Ile Leu Leu Ala Phe Trp Met Tyr Arg His
          145          150          155          160
Arg Lys Pro Pro Tyr Gly His Val Glu Ile Asn Glu Asp Pro Gly Leu
          165          170          175
Pro Pro Pro Ser Pro Leu Val Gly Leu Lys Pro Leu Gln Leu Leu Glu
          180          185          190
Ile Lys Ala Arg Gly Arg Phe Gly Cys Val Trp Lys Ala Arg Leu Leu
          195          200          205
Asn Glu Tyr Val Ala Val Lys Ile Phe Pro Val Gln Asp Lys Gln Ser
          210          215          220
Trp Gln Cys Glu Lys Glu Ile Phe Thr Thr Pro Gly Met Lys His Glu
          225          230          235          240
Asn Leu Leu Glu Phe Ile Ala Ala Glu Lys Arg Gly Ser Asn Leu Glu
          245          250          255
Met Glu Leu Trp Leu Ile Thr Ala Phe His Asp Lys Gly Ser Leu Thr
          260          265          270
Asp Tyr Leu Lys Gly Asn Leu Val Ser Trp Asn Glu Leu Cys His Ile
          275          280          285
Thr Glu Thr Met Ala Arg Gly Leu Ala Tyr Leu His Glu Asp Val Pro
          290          295          300

```

Arg	Cys	Lys	Gly	Glu	Gly	His	Lys	Pro	Ala	Ile	Ala	His	Arg	Asp	Phe
305					310					315					320
Lys	Ser	Lys	Asn	Val	Leu	Leu	Arg	Asn	Asp	Leu	Thr	Ala	Ile	Leu	Ala
			325						330					335	
Asp	Phe	Gly	Leu	Ala	Val	Arg	Phe	Glu	Pro	Gly	Lys	Pro	Pro	Gly	Asp
			340					345					350		
Thr	His	Gly	Gln	Val	Gly	Thr	Arg	Arg	Tyr	Met	Ala	Pro	Glu	Val	Leu
		355					360					365			
Glu	Gly	Ala	Ile	Asn	Phe	Gln	Arg	Asp	Ser	Phe	Leu	Arg	Ile	Asp	Met
	370					375					380				
Tyr	Ala	Met	Gly	Leu	Val	Leu	Trp	Glu	Ile	Val	Ser	Arg	Cys	Thr	Ala
385					390					395					400
Ala	Asp	Gly	Pro	Val	Asp	Glu	Tyr	Leu	Leu	Pro	Phe	Glu	Glu	Glu	Ile
				405					410					415	
Gly	Gln	His	Pro	Ser	Leu	Glu	Asp	Leu	Gln	Glu	Val	Val	Val	His	Lys
			420					425						430	
Lys	Ile	Arg	Pro	Val	Phe	Lys	Asp	His	Trp	Leu	Lys	His	Pro	Gly	Leu
		435					440					445			
Ala	Gln	Leu	Cys	Val	Thr	Ile	Glu	Glu	Cys	Trp	Asp	His	Asp	Ala	Glu
	450					455					460				
Ala	Arg	Leu	Ser	Ala	Gly	Cys	Val	Glu	Glu	Arg	Ile	Ser	Gln	Ile	Arg
465					470					475					480
Lys	Ser	Val	Asn	Gly	Thr	Thr	Ser	Asp	Cys	Leu	Val	Ser	Ile	Val	Thr
				485					490					495	
Ser	Val	Thr	Asn	Val	Asp	Leu	Pro	Pro	Lys	Glu	Ser	Ser	Ile		
			500					505					510		

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 6 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: both

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Asp Leu Lys Pro Glu Asn
1 5

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 6 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: both

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(ix) FEATURE:
 (A) NAME/KEY: Modified-site
 (B) LOCATION: 2
 (D) OTHER INFORMATION: /note= "Xaa at position 2 is either
 "Thr" or "Ser"."

(ix) FEATURE:
 (A) NAME/KEY: Modified-site
 (B) LOCATION: 5
 (D) OTHER INFORMATION: /note= "Xaa a position 5 is either
 "Tyr" or "Phe"."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Gly Xaa Xaa Xaa Xaa Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: both

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Asp Ile Lys Ser Lys Asn
1 5

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: both

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Gly Thr Arg Arg Tyr Met
1 5

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Asp Leu Ala Ala Arg Asn
1 5

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: both

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 3

(D) OTHER INFORMATION: /note= "Xaa at position 3 is either
"Ile" or "Val"."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 4

(D) OTHER INFORMATION: /note= "Xaa at position 4 is either
"Lys" or "Arg"."

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 6

(D) OTHER INFORMATION: /note= "Xaa at position 6 is either
"Thr" or "Met"."

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Xaa Pro Xaa Xaa Trp Xaa
1 5

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1602 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: both

(D) TOPOLOGY: both

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 72..1553

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
CCCCGGAAC TTCAAAGCGC GCTGCGGCGG CGCTCTGGGA CCCCAGGCC TTGCACCGCC      60
GCGGGGTGGC C ATG ACC CCA GCG CGC CGC TCC GCA CTG AGC CTG GCC CTC      110
          Met Thr Pro Ala Arg Ser Ala Leu Ser Leu Ala Leu
          1          5          10
```

CTG Leu	CTG Leu	GTG Val	GCA Ala	CTG Leu	GCC Ala	TCC Ser	GAC Asp	CTT Leu	GCG Ala	GCA Ala	GGA Gly	CTG Leu	AAG Lys	TGT Cys	GTG Val	158
1520																
TGT Cys	CTT Leu	TTG Leu	TGT Cys	GAT Asp	TCC Ser	TCA Ser	AAC Asn	TTT Phe	ACC Thr	TGC Cys	CAA Gln	ACC Thr	GAA Glu	GGA Gly	GCA Ala	206
303545																
TGC Cys	TGG Trp	GCC Ala	TCT Ser	GTC Val	ATG Met	CTA Leu	ACC Thr	AAC Asn	GGG Gly	AAA Lys	GAA Glu	CAG Gln	GTG Val	AGC Ser	AAA Lys	254
5060																
TCG Ser	TGC Cys	GTG Val	TCC Ser	CTC Leu	CCG Pro	GAA Glu	CTA Leu	AAT Asn	GCT Ala	CAG Gln	GTC Val	TTC Phe	TGT Cys	CAC His	AGT Ser	302
70																
TCC Ser	AAC Asn	AAC Asn	GTG Val	ACC Thr	AAG Lys	ACC Thr	GAA Glu	TGT Cys	TGC Cys	TTC Phe	ACA Thr	GAC Asp	TTC Phe	TGC Cys	AAC Asn	350
808590																
AAC Asn	ATC Ile	ACT Thr	CAG Gln	CAC His	CTT Leu	CCC Pro	ACA Thr	GCA Ala	TCT Ser	CCA Pro	GAT Asp	GCC Ala	CCT Pro	AGA Arg	CTT Leu	398
95100105																
GGC Gly	CCC Pro	ACA Thr	GAG Glu	CTG Leu	ACA Thr	GTT Val	GTT Val	ATC Ile	ACT Thr	GTA Val	CCT Pro	GTT Val	TGC Cys	CTC Leu	CTG Leu	446
110115120125																
TCC Ser	ATC Ile	GCA Ala	GCC Ala	ATG Met	CTA Leu	ACG Thr	ATA Ile	TGG Trp	GCC Ala	TGC Cys	CAG Gln	GAC Asp	CGC Arg	CAG Gln	TGC Cys	494
130135140																
ACA Thr	TAC Tyr	AGG Arg	AAG Lys	ACC Thr	AAG Lys	AGA Arg	CAC His	AAT Asn	GTG Val	GAG Glu	GAA Glu	CCA Pro	CTG Leu	GCA Ala	GAG Glu	542
145150155																
TAC Tyr	AGC Ser	CTT Leu	GTC Val	AAT Asn	GCT Ala	GGA Gly	AAA Lys	ACC Thr	CTC Leu	AAA Lys	GAT Asp	CTG Leu	ATT Ile	TAT Tyr	GAT Asp	590
160165170																
GCC Ala	ACT Thr	GCC Ala	TCG Ser	GGC Gly	TCA Ser	GGA Gly	TCT Ser	GGC Gly	CCG Pro	CCT Pro	CTT Leu	TTG Leu	GTT Val	CAA Gln	AGA Arg	638
175180185																
ACC Thr	ATC Ile	GCA Ala	AGG Arg	ACA Thr	ATT Ile	GTA Val	CTT Leu	CAA Gln	GAA Glu	ATC Ile	GTA Val	GGA Gly	AAA Lys	GGT Gly	CGG Arg	686
190195200205																
TTT Phe	GGG Gly	GAA Glu	GTG Val	TGG Trp	CAC His	GGA Gly	AGA Arg	TGG Trp	TGT Cys	GGA Gly	GAA Glu	GAT Asp	GTG Val	GCT Ala	GTG Val	734
210215220																
AAA Lys	ATA Ile	TTC Phe	TCC Ser	TCC Ser	AGA Arg	GAT Asp	GAG Glu	AGA Arg	TCT Ser	TGG Trp	TTC Phe	CGT Arg	GAG Glu	GCA Ala	GAA Glu	782
225230235																
ATT Ile	TAT Tyr	CAG Gln	ACG Thr	GTA Val	ATG Met	CTG Leu	AGA Arg	CAT His	GAG Glu	AAT Asn	ATT Ile	CTC Leu	GGT Gly	TTC Phe	ATC Ile	830
240245250																
GCG Ala	GCC Ala	GAC Asp	AAC Asn	AAA Lys	GAT Asp	AAT Asn	GGA Gly	ACC Thr	TGG Trp	ACT Thr	CAG Gln	CTT Leu	TGG Trp	CTT Leu	GTG Val	878
255260265																
TCA Ser	GAG Glu	TAT Tyr	CAC His	GAG Glu	CAG Gln	GGC Gly	TCC Ser	TTA Leu	TAT Tyr	GAC Asp	TAT Tyr	TTG Leu	AAT Asn	AGA Arg	AAC Asn	270275280285

ATA	GTG	ACC	GTG	GCT	GGA	ATG	GTC	AAG	TTG	GCG	CTT	TCA	ATA	GCG	AGT		974
Ile	Val	Thr	Val	Ala	Gly	Met	Val	Lys	Leu	Ala	Leu	Ser	Ile	Ala	Ser		
				290					295					300			
GGT	CTG	GCT	CAC	CTA	CAC	ATG	GAG	ATC	GTG	GGC	ACT	CAA	GGT	AAG	CCT		1022
Gly	Leu	Ala	His	Leu	His	Met	Glu	Ile	Val	Gly	Thr	Gln	Gly	Lys	Pro		
			305					310					315				
GCT	ATT	GCT	CAC	CGA	GAT	ATA	AAG	TCA	AAG	AAT	ATC	TTA	GTC	AAA	AAG		1070
Ala	Ile	Ala	His	Arg	Asp	Ile	Lys	Ser	Lys	Asn	Ile	Leu	Val	Lys	Lys		
		320					325					330					
TGT	GAC	ACT	TGT	GCC	ATA	GCT	GAC	TTA	GGG	CTG	GCT	GTG	AAA	CAT	GAT		1118
Cys	Asp	Thr	Cys	Ala	Ile	Ala	Asp	Leu	Gly	Leu	Ala	Val	Lys	His	Asp		
	335					340					345						
TCT	ATC	ATG	AAC	ACT	ATA	GAT	ATA	CCC	CAG	AAT	CCT	AAA	GTG	GGA	ACC		1166
Ser	Ile	Met	Asn	Thr	Ile	Asp	Ile	Pro	Gln	Asn	Pro	Lys	Val	Gly	Thr		
350					355				360						365		
AAG	AGG	TAT	ATG	GCT	CCC	GAA	ATG	CTT	GAT	GAT	ACA	ATG	AAC	GTC	AAC		1214
Lys	Arg	Tyr	Met	Ala	Pro	Glu	Met	Leu	Asp	Asp	Thr	Met	Asn	Val	Asn		
				370					375					380			
ATC	TTT	GAG	TCC	TTC	AAG	CGA	GCT	GAC	ATC	TAT	TCG	GTG	GGG	CTG	GTT		1262
Ile	Phe	Glu	Ser	Phe	Lys	Arg	Ala	Asp	Ile	Tyr	Ser	Val	Gly	Leu	Val		
			385					390					395				
TAC	TGG	GAA	ATA	GCT	CGA	AGG	TGT	TCA	GTT	GGA	GGA	CTT	GTT	GAA	GAG		1310
Tyr	Trp	Glu	Ile	Ala	Arg	Arg	Cys	Ser	Val	Gly	Gly	Leu	Val	Glu	Glu		
		400					405					410					
TAC	CAG	TTG	CCT	TAT	TAT	GAC	ATG	GTG	CCT	TCA	GAT	CCT	TCC	ATA	GAG		1358
Tyr	Gln	Leu	Pro	Tyr	Tyr	Asp	Met	Val	Pro	Ser	Asp	Pro	Ser	Ile	Glu		
	415					420					425						
GAA	ATG	AGG	AAG	GTC	GTT	TGT	GAT	CAG	AAA	CTG	CGA	CCA	AAT	CTC	CCA		1406
Glu	Met	Arg	Lys	Val	Val	Cys	Asp	Gln	Lys	Leu	Arg	Pro	Asn	Leu	Pro		
430					435				440						445		
AAC	CAG	TGG	CAA	AGC	TGT	GAG	GCG	CTC	CGG	GTC	ATG	GGA	AGA	ATA	ATG		1454
Asn	Gln	Trp	Gln	Ser	Cys	Glu	Ala	Leu	Arg	Val	Met	Gly	Arg	Ile	Met		
				450					455					460			
CGT	GAG	TGC	TGG	TAT	GCC	AAC	GGG	GCA	GCT	CGC	CTG	ACC	GCC	CTG	CGC		1502
Arg	Glu	Cys	Trp	Tyr	Ala	Asn	Gly	Ala	Ala	Arg	Leu	Thr	Ala	Leu	Arg		
			465					470					475				
GTG	AAG	AAG	ACC	ATT	TCT	CAG	CTG	TGT	GTC	AAG	GAA	GAC	TGT	AAG	GCC		1550
Val	Lys	Lys	Thr	Ile	Ser	Gln	Leu	Cys	Val	Lys	Glu	Asp	Cys	Lys	Ala		

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 493 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEO ID NO:12:

Met 1	Thr	Pro	Ala	Arg 5	Arg	Ser	Ala	Leu	Ser 10	Leu	Ala	Leu	Leu	Leu	Val
Ala	Leu	Ala	Ser 20	Asp	Leu	Ala	Ala	Gly 25	Leu	Lys	Cys	Val	Cys 30	Leu	Leu
Cys	Asp	Ser 35	Ser	Asn	Phe	Thr	Cys 40	Gln	Thr	Glu	Gly	Ala 45	Cys	Trp	Ala
Ser	Val 50	Met	Leu	Thr	Asn	Gly 55	Lys	Glu	Gln	Val	Ser 60	Lys	Ser	Cys	Val
Ser 65	Leu	Pro	Glu	Leu	Asn 70	Ala	Gln	Val	Phe	Cys 75	His	Ser	Ser	Asn	Asn 80
Val	Thr	Lys	Thr	Glu 85	Cys	Cys	Phe	Thr	Asp 90	Phe	Cys	Asn	Asn	Ile 95	Thr
Gln	His	Leu	Pro 100	Thr	Ala	Ser	Pro	Asp 105	Ala	Pro	Arg	Leu	Gly 110	Pro	Thr
Glu	Leu	Thr 115	Val	Val	Ile	Thr	Val 120	Pro	Val	Cys	Leu 125	Leu	Ser	Ile	Ala
Ala	Met 130	Leu	Thr	Ile	Trp	Ala 135	Cys	Gln	Asp	Arg	Gln 140	Cys	Thr	Tyr	Arg
Lys 145	Thr	Lys	Arg	His 150	Asn	Val	Glu	Glu	Pro	Leu 155	Ala	Glu	Tyr	Ser	Leu 160
Val	Asn	Ala	Gly 165	Lys	Thr	Leu	Lys	Asp 170	Leu	Ile	Tyr	Asp	Ala	Thr 175	Ala
Ser	Gly	Ser 180	Gly	Ser	Gly	Pro	Pro	Leu 185	Leu	Val	Gln	Arg	Thr 190	Ile	Ala
Arg	Thr	Ile 195	Val	Leu	Gln	Glu	Ile 200	Val	Gly	Lys	Gly	Arg 205	Phe	Gly	Glu
Val	Trp 210	His	Gly	Arg	Trp	Cys 215	Gly	Glu	Asp	Val	Ala 220	Val	Lys	Ile	Phe
Ser 225	Ser	Arg	Asp	Glu	Arg 230	Ser	Trp	Phe	Arg	Glu 235	Ala	Glu	Ile	Tyr	Gln 240
Thr	Val	Met	Leu	Arg 245	His	Glu	Asn	Ile	Leu 250	Gly	Phe	Ile	Ala	Ala 255	Asp
Asn	Lys	Asp	Asn 260	Gly	Thr	Trp	Thr	Gln 265	Leu	Trp	Leu	Val	Ser 270	Glu	Tyr
His	Glu	Gln 275	Gly	Ser	Leu	Tyr	Asp 280	Tyr	Leu	Asn	Arg	Asn 285	Ile	Val	Thr
Val	Ala 290	Gly	Met	Val	Lys	Leu 295	Ala	Leu	Ser	Ile	Ala 300	Ser	Gly	Leu	Ala

(2) INFORMATION FOR SEQ ID NO:13:

(ii) MOLECULE TYPE: DNA (genomic)

CGGGATCCGT NGCNGTNAAR ATHTTYCC

28

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 35 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CGGGATCCYT CNGGNGCCAT RTANCKYCTN GTNCC

35